AD-776 609

THE VALUE OF SEQUENTIAL INFORMATION

Allen Clinton Miller, III

Stanford University

Prepared for:

Office of Naval Research
Advanced Research Projects Agency
National Science Foundation

31 January 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

AD776609

Research Report No. EES-DA-73-1
January 1974

THE VALUE OF SEQUENTIAL INFORMATION

ALLEN C. MILLER III

DECISION ANALYSIS PROGRAM
Professor Ronald A. Howard
Principal Investigator

DEPARTMENT OF ENGINEERING-ECONOMIC SYSTEMS
Stanford University
Stanford, California 94305

SPONSORSHIPS

Advanced Research Projects Agency, Human Resources Research Office ARPA Order No. 2449, monitored by Engineering Psychology Programs, Office of Naval Research, under Contract No. N00014-67-A-0112-0077 (NR 197-024) covering the period May 1, 1973 to December 30, 1973.

National Science Foundation under NSF Grant GK-36491.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER EES-DA-73-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) "The Value of Sequential Information"		5. TYPE OF REPORT & PERIOD COVERED Technical 5/1/73 to 12/30/73
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Allen Clinton Miller, III		8. CONTRACT OR GRANT NUMBER(s) N00014-67-A-0112-0077
9. PERFORMING ORGANIZATION NAME AND ADDRESS The Board of Trustees of the Leland Stanford Junior University, c/o Office of Research Administrator, Encina Hall, Stanford, California 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 000000 ARPA ORDER #2449
11. CONTROLLING OFFICE NAME AND ADDRESS Advanced Research Projects Agency Human Resources Research Office 1400 Wilson Blvd., Arlington, Virginia 22209		12. REPORT DATE January 31, 1974
		13. NUMBER OF PAGES 205 pages
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Engineering Psychology Programs, Code 455 Office of Naval Research 800 North Quincy Street; Arlington, Virginia 22217		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

A dissertation submitted to the Department of Engineering-Economic Systems and the Committee on Graduate Studies of Stanford University in partial fulfillment of the Requirements for the Degree of Doctor of Philosophy, November 1973.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

DECISION ANALYSIS	DECISION THEORY	SEQUENTIAL INFORMATION
DECISION-MAKING	INFORMATION	VALUE OF INFORMATION

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

In decision analysis we normally consider the value of information to be a constant against which the cost of information is compared. However, when it is possible to buy information sequentially, the value of information is not a constant. Rather, it is a function of the prices of the various pieces of information, or "observables." When we are faced with a decision and learn one observable, this information not only helps us make the original decision, but also helps us decide if we should pay for more observables. For this

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Block 20 (continued)

reason, the first observable has a value above and beyond that which we would assign if there were no possibility of obtaining additional information.

As the cost of one observable is increased, the value of information about another observable can decrease or remain constant. In fact, the derivative of the value of information about one observable with respect to the cost of another observable must lie in the closed interval [-1,0] . The second derivative of the value of information with respect to the price of any observable cannot be negative.

Since the value of information is a function of the prices of the observables, it is necessary to know the prices of all the observables before deciding whether or not to pay for a single piece of information. If we know the price of every observable, we can determine which, if any, we should buy first. In this manner, we can divide the set of all possible n-tuples of observable prices into mutally-exclusive and collectively-exhaustive subsets such that our best initial decision is to buy the i^{th} observable when the n-tuple of prices is contained in the i^{th} subset. The subsets can be viewed as regions in the n-dimensional Euclidean space spanned by the prices; we can approximate or bound the regions.

If the prices of the observables are uncertain, the value of information is a function of our state of information about the prices. If we assume that the prices are independent random variables, the value of information is a function of the expected prices with the same functional form that would result if the prices were certain. If the cost of learning an observable changes after other observables are purchased, the value of information then depends on all of the possible prices.

ABSTRACT

In decision analysis we normally consider the value of information to be a constant against which the cost of information is compared. However, when it is possible to buy information sequentially, the value of information is not a constant. Rather, it is a function of the prices of the various pieces of information, or "observables." When we are faced with a decision and learn one observable, this information not only helps us make the original decision, but also helps us decide if we should pay for more observables. For this reason, the first observable has a value above and beyond that which we would assign if there were no possibility of obtaining additional information.

As the cost of one observable is increased, the value of information about another observable can decrease or remain constant. In fact, the derivative of the value of information about one observable with respect to the cost of another observable must lie in the closed interval [-1,0] . The second derivative of the value of information with respect to the price of any observable cannot be negative.

Since the value of information is a function of the prices of the observables, it is necessary to know the prices of all the observables before deciding whether or not to pay for a single piece of information. If we know the price of every observable, we can determine which, if any, we should buy first. In this manner, we can divide the set of all possible n-tuples of observable prices into mutually-exclusive and collectively-exhaustive subsets such that our best initial decision is to buy the i^{th} observable when the n-tuple of prices is contained in

the i^{th} subset. The subsets can be viewed as regions in the n-dimensional Euclidean space spanned by the prices; we can approximate or bound the regions.

If the prices of the observables are uncertain, the value of information is a function of our state of information about the prices. If we assume that the prices are independent random variables, the value of information is a function of the expected prices with the same functional form that would result if the prices were certain. If the cost of learning an observable changes after other observables are purchased, the value of information then depends on all of the possible prices.

ACKNOWLEDGEMENTS

I would like to express my appreciation to Professor Ronald A. Howard for all of his advice and encouragement. In addition to guiding the research that led to this dissertation, he taught the courses that originally motivated my interest in decision analysis.

Thanks are due to Dr. James E. Matheson and Professor Donald A. Dunn for reviewing and commenting on the manuscript. I would also like to thank Professor Richard D. Smallwood for his comments on the content of the dissertation.

My friends and fellow students in the Department of Engineering-Economic Systems assisted me with numerous discussions of topics related to my dissertation. I owe them all a debt of gratitude. I would especially like to thank Verne G. Chant for his detailed review on an early draft of my dissertation.

I am grateful to my wife, Susan, for her many readings of the manuscript, and for her encouragement and patience.

I would also like to thank Mrs. Ditter Peschcke-Koedt for typing the manuscript, and Mrs. Louise Goodrich for helping me with all of the necessary secretarial chores.

Finally I am grateful to the National Science Foundation for providing me with financial support, in the form of an NSF Traineeship, during the time that I was working on this dissertation.

This research was partially supported by the Advanced Research Projects Agency of the Department of Defense, as monitored by the Office of Naval Research under Contract No. N00014-67-A-0112-0077. This research was also partially supported by the National Science Foundation under NSF Grant GK-36491.

TABLE OF CONTENTS

LIST OF FIGURES

Figure		Page

Figure		Page

CHAPTER 1

INTRODUCTION

In decision analysis, [6,11,12,13],* we customarily think of the value of a piece of information as being a fixed amount against which we should compare the cost of learning the information. We should refuse to pay for the information if its cost exceeds its value. Otherwise we can expect to gain, in some statistical sense, by paying for the information. It is possible to extend this simple idea by incorporating the concepts of risk preference, utility functions, and imperfect information. However the basic concept remains the same. In principle we can determine a fixed maximum price that we should pay for a piece of information, ~~and~~ then use this quantity, which we call the value of the information, to decide whether or not to buy the information.

Conclusions and Contributions

Unfortunately, as the following chapters will show, this straightforward notion of the value of information is inadequate. When we have the option of buying a piece of information--which we call an "observable"--and studying it before deciding whether or not to pay for additional information, the value of the information is greater than it would be if we could buy only the observable by itself. When information is available sequentially, we must consider the prices of all of the observables in order to decide whether or not to pay for any individual

---

*Numbers in square brackets refer to the Bibliography.

observable. In other words, the value of information, rather than being a constant, is a function of the prices of all of the other observables that we might decide to buy.

This idea, which will be explored in some detail in the following chapters, is one of the general conclusions and contributions of this paper. We can state this conclusion more precisely with the following definitions. If there are n observables that we could learn sequentially, then we can define n random variables $(y_1, \ldots, y_n)$ such that learning the i^{th} piece of information is equivalent to learning the actual value of y_i. Since learning y_i is equivalent to learning the i^{th} piece of information, we shall call y_i the i^{th} observable. Let K_{y_i} and V_{y_i} be the cost and value, respectively, of the i^{th} observable. Using these definitions, the conclusion stated above becomes:

1. In general, V_{y_i} is a function of the prices of all of the observables except K_{y_i}. Thus

$$V_{y_i} = V_{y_i}(K_{y_1}, \ldots, K_{y_{i-1}}, K_{y_{i+1}}, \ldots, K_{y_n})$$

For simplicity we sometimes write the extected value of sequential information about y_i as $V_{y_i}(K_{y_1}, \ldots, K_{y_n})$, where it is understood that V_{y_i} does not depend on K_{y_i}. Since our decisions to pay for or refuse the information represented by y_i depends on the relative values of K_{y_i} and V_{y_i}, we will have to know all of the prices including K_{y_i} before we can make the decision. The method used to determine the value function $V_{y_i}(K_{y_1}, \ldots, K_{y_n})$ is demonstrated in the following chapters. If the value of a piece of information is a function of the prices

of all of the other observables, what can we say about the form of this function? If the decision maker's utility function has certain properties, we can show that:

2. Increasing any one of the prices of the observables will cause the value of a piece of information to remain constant or decrease. However the value of the information cannot decrease by more than the amount that the price increases. In other words,

$$\frac{\partial V_{y_i}}{\partial K_{y_j}} (K_{y_1}, \ldots, K_{y_{i-1}}, K_{y_{i+1}}, \ldots, K_{y_n}) \in [-1,0]$$

This result, and those that follow, hold for any decision maker who has a utility function that satisfies the delta property and is monotonically increasing (the delta property is discussed in the next section). It is possible for $\partial V_{y_i}/\partial K_{y_j}$ to equal any value between -1 and 0, so that increasing a price by one unit can cause the value of some observables to drop by a fraction of a unit.

We can go one step further and show that $\partial V_{y_i}/\partial K_{y_j}$ must increase or remain constant when K_{y_j} increases:

3. Increasing any one of the prices will cause the rate of change of the value of any observable with respect to that price to increase or remain constant. In other words,

$$\frac{\partial^2 V_y}{\partial K_{y_j}^2} (K_{y_1}, \ldots, K_{y_{i-1}}, K_{y_{i+1}}, \ldots, K_{y_n}) \geq 0$$

It is possible for the first derivative to be discontinuous,

and for the second derivative to be infinite. However, it need not be true that

$$\frac{\partial^2 V_{y_i}}{\partial K_{y_j} \partial K_{y_k}} \geq 0$$

It is possible for $\partial V_{y_i}/\partial K_{y_j}$ to make a discontinuous jump to a lower value as K_{y_k} is increased.

We can write a general expression for the value function $V_{y_i}(K_{y_1}, \ldots, K_{y_n})$ as shown in Chapter 4. However, this expression is not useful for solving practical problems since it is written as a series of maximizations. For any given problem we can carry out a number of calculations to determine algebraic expressions for the value functions that do not involve maximization. Using these expressions we can determine which observable, if any, we should buy first for any given set of prices, $(K_{y_1}, \ldots, K_{y_n})$. In this manner we can divide the set of all possible n-tuples of prices into subsets, such that when the n-tuple of prices is in the i^{th} subset, our best initial decision is to buy the i^{th} observable. Thus:

4. It is possible to determine $(n+1)$ mutually exclusive and collectively exhaustive sets Ω_i for $i = 0,1,\ldots,n$, such that when

$$(K_{y_1}, \ldots, K_{y_n}) \in \Omega_i$$

our best initial decision is to buy the i^{th} piece of information. If i equals zero, our best decision is to refuse all the information. We can visualize the Ω_i as decision regions in the n-dimensional Euclidean space spanned by the prices $(K_{y_1}, \ldots, K_{y_n})$.

Unfortunately a large number of calculations are required to determine the value functions when there are more than two or three pieces of sequential information available. It will be shown in Chapter 4 that the number of calculations increases like 2^n, where n is the number of sequential pieces of information. However, even when the number of calculations required to determine the value functions and the decision sets, Ω_i, become prohibitively large, we can determine certain properties of these functions and sets by carrying out a relatively small number of calculations. In particular:

5. It is possible to approximate the decision sets by solving the appropriate decision tree using several sets of prices for the observables. Also, there exists a maximum price that we are willing to pay for each observable when our best initial decision is to buy that observable. It is possible to bound this maximum price.

One bound that works for all sequential information problems is the value of learning simultaneously all of the observables that are available. This constant is relatively easy to calculate and it bounds the maximum initial value of every piece of information. For certain problems it is possible to find tighter bounds with a little more effort; the procedures for doing so will be discussed in the following chapters.

Utility and Risk Preference

We assume throughout this paper that the decision maker always prefers more money to less money. In other words, his utility function [5,9,10,15] must be monotonically increasing. The conclusions stated previously are based on this assumption.

We have defined the value of information as the maximum price that we are willing to pay for the information. It is difficult to determine this maximum price unless the decision maker's utility function has certain properties. For an arbitrary utility function it is necessary to compare the expected utilities that result when we learn and do not learn the information, and then vary the price of the information until the two expected utilities are the same. For simple problems, it is possible to solve an equation for the maximum price, but for more difficult problems the result is often determined by trial and error.

This difficulty can be overcome if the decision maker's risk preference can be represented by a utility function that satisfies the "delta property." The delta property implies that adding a given amount to the value of all of the possible outcomes associated with any lottery increases the certain equivalent of the lottery by exactly the same amount. The certain equivalent is that value whose utility is equal to the expected utility of the lottery. Mathematically the delta property requires that the utility function $U(.)$ satisfy

$$U(\tilde{x} + \Delta) = \alpha U(x_1 + \Delta) + (1 - \alpha)\, U(x_2 + \Delta)$$

whenever

$$U(\tilde{x}) = \alpha U(x_1) + (1 - \alpha)\, U(x_2)$$

$\tilde{x}$ is the certain equivalent of this lottery. This must hold for all x_1, x_2, and Δ, and all $\alpha \varepsilon [0,1]$. It can be shown, [5,10], that all utility functions that satisfy the delta property have one of the following forms:

$$U(x) = A + Bx$$

$$U(x) = A + Be^{-\gamma x}$$

The first of these utility functions, with $A = 0$ and $B = 1$, corresponds to a decision maker who attempts to maximize his expected profit in an uncertain situation. We call such a person an expected-profit or risk-indifferent decision maker. (As long as B is positive, the first utility function can always be used to represent an expected-profit decision maker. Maximizing the first utility function with B positive is equivalent to maximizing expected profit.) The second utility function can represent a risk-averse or risk-preferring decision maker.

With the exception of the last part of Chapter 4, the following discussion is presented in terms of an expected-profit decision maker. However, it will be shown that the general conclusions stated previously are valid as long as the decision maker has a utility function that satisfies the delta property and is monotonically increasing.

When the utility function satisfies the delta property, the calculations required to determine the value functions $V_{y_i}(K_{y_1}, \ldots, K_{y_n})$ are not significantly longer or more difficult than those required for an expected-profit decision maker. In this case the value of a piece of information can be determined by taking the difference of certain equivalents. On the other hand, when the utility function does not satisfy the delta property, the calculations required to determine the value functions become very long and difficult.

Individual, Simultaneous, and Sequential Information

When there is more than one piece of information that we can learn, we can have the option of learning the information individually, simultaneously, or sequentially. As we shall see, a piece of information can have different values depending on how we decide to learn it. For this

reason we represent the values of individual, simultaneous, and sequential information with different symbols.

If we decide to learn an individual piece of information and refuse all of the remaining information, the maximum price that we are willing to pay is $V^N_{y_i}$, where y_i is the observable we are learning. $V^N_{y_i}$ is called the value of individual information about the observable y_i. In this work a value of information with a superscript N means that the value is determined by assuming that no additional information will be purchased beyond that shown in the subscript.

On the other hand we could decide to learn several pieces of information simultaneously. In this case we would receive all of the information at once without an opportunity to look at one observable before deciding whether or not to pay for another. If we learn two observables simultaneously, the expected value of learning both is $V^N_{y_i y_j}$, where y_i and y_j are the observables we are learning. $V^N_{y_i y_j}$ is called the value of simultaneous information about y_i and y_j. If there are three pieces of information, the expected value of learning all three simultaneously is $V^N_{y_i y_j y_k}$, and so on. Again, the superscript N means that no additional information will be purchased.

$V^N_{y_i}$ and $V^N_{y_i y_j}$ are constants that do not depend on the cost of any observable. We might expect the value of learning two pieces of information simultaneously to equal the sum of the values of learning them individually. In other words, we might expect $V^N_{y_i y_j}$ to equal the sum of $V^N_{y_i}$ and $V^N_{y_j}$. However, this is not generally true. It is possible for $V^N_{y_i y_j}$ to be greater or less than the sum of $V^N_{y_i}$ and $V^N_{y_j}$. (Similar statements can be made when there are more than two observables. For example, it is possible for $V^N_{y_i y_j y_k}$ to be greater or less than the sum of $V^N_{y_i}$, $V^N_{y_j}$, and $V^N_{y_k}$.) This can be demonstrated with two simple decision problems.

Decision Problem One: Suppose we have an opportunity to play the following game. A fair coin will be flipped twice, and we will win ten dollars if we can correctly predict whether or not the two outcomes will be the same. It is easy to see that our expected profit in the absence of any information about the outcomes of the two flips is five dollars. Now suppose that a reliable clairvoyant offers to tell us the outcome of either or both flips. How much is his information worth to us if we are trying to maximize our expected profit?

We have an opportunity to learn two different pieces of information, so there are two observables in this problem. Let y_i $(i = 1,2)$ be a random variable such that y_i equals zero if the outcome of the i^{th} flip is a tail, and y_i equals one if the outcome of the i^{th} flip is a head. If we decide to learn y_1 only, our expected profit after receiving the information is still five dollars. Since learning y_1 does not increase our expected profit, $V^N_{y_1}$ equals zero. Similarly, $V^N_{y_2}$ equals zero. On the other hand, if we decide to learn both y_1 and y_2, we will know whether or not the two outcomes are the same and our expected profit will be ten dollars. Since learning both observables increases our expected profit by five dollars, $V^N_{y_1 y_2}$ equals five dollars. Thus $V^N_{y_1 y_2}$ exceeds the sum of $V^N_{y_1}$ and $V^N_{y_2}$ for this problem.

Decision Problem Two: Suppose we have an opportunity to play the following game. We make a single prediction of either heads or tails, and then a fair coin is flipped twice. We are given ten dollars each time the outcome of a flip agrees with our prediction. However, we cannot predict heads on one flip and tails on the other. In the absence of any information about the outcomes of the two flips, our expected profit is ten dollars. Now suppose that the reliable clairvoyant again offers to tell us

the outcome of either or both flips. How much is his information worth this time?

Let y_1 and y_2 be the two observables, as in the previous decision problem. If we decide to learn y_1 only, we will predict the outcome that the clairvoyant says will occur. This means that we are certain to receive ten dollars for correctly predicting the first flip, and we have an expected profit of five dollars from the second flip. Thus our total expected profit increases to fifteen dollars, and $V_{y_1}^N$ equals five dollars. Similarly, $V_{y_2}^N$ equals five dollars. On the other hand, if we decide to learn both y_1 and y_2, we will be able to use this information to predict both flips correctly and collect twenty dollars with a probability of 50%. Otherwise the clairvoyant will tell us that the two outcomes are different, and we will only be able to make ten dollars. In this case our expected profit is fifteen dollars, and $V_{y_1y_2}^N$ equals five dollars. Thus $V_{y_1y_2}^N$ is less than the sum of $V_{y_1}^N$ and $V_{y_2}^N$ for this decision problem.

These simple examples show that our intuition does not lead us to the correct conclusion. The value of learning several pieces of information simultaneously need not equal the sum of the values of learning each piece of information individually. LaValle [7,8] describes this situation as being "roughly analogous to saying that the whole is greater than the sum of its parts."

We can understand this situation better by graphing the prices for which we would be willing to buy one or both pieces of information in the two simple decision problems. Let K_{y_1} and K_{y_2} be the cost of learning the actual values of y_1 and y_2 respectively. Any pair of prices (K_{y_1}, K_{y_2}) can be represented by a point in a two-dimensional diagram such as those shown in Figs. 1.1 and 1.2. We are willing to buy the information

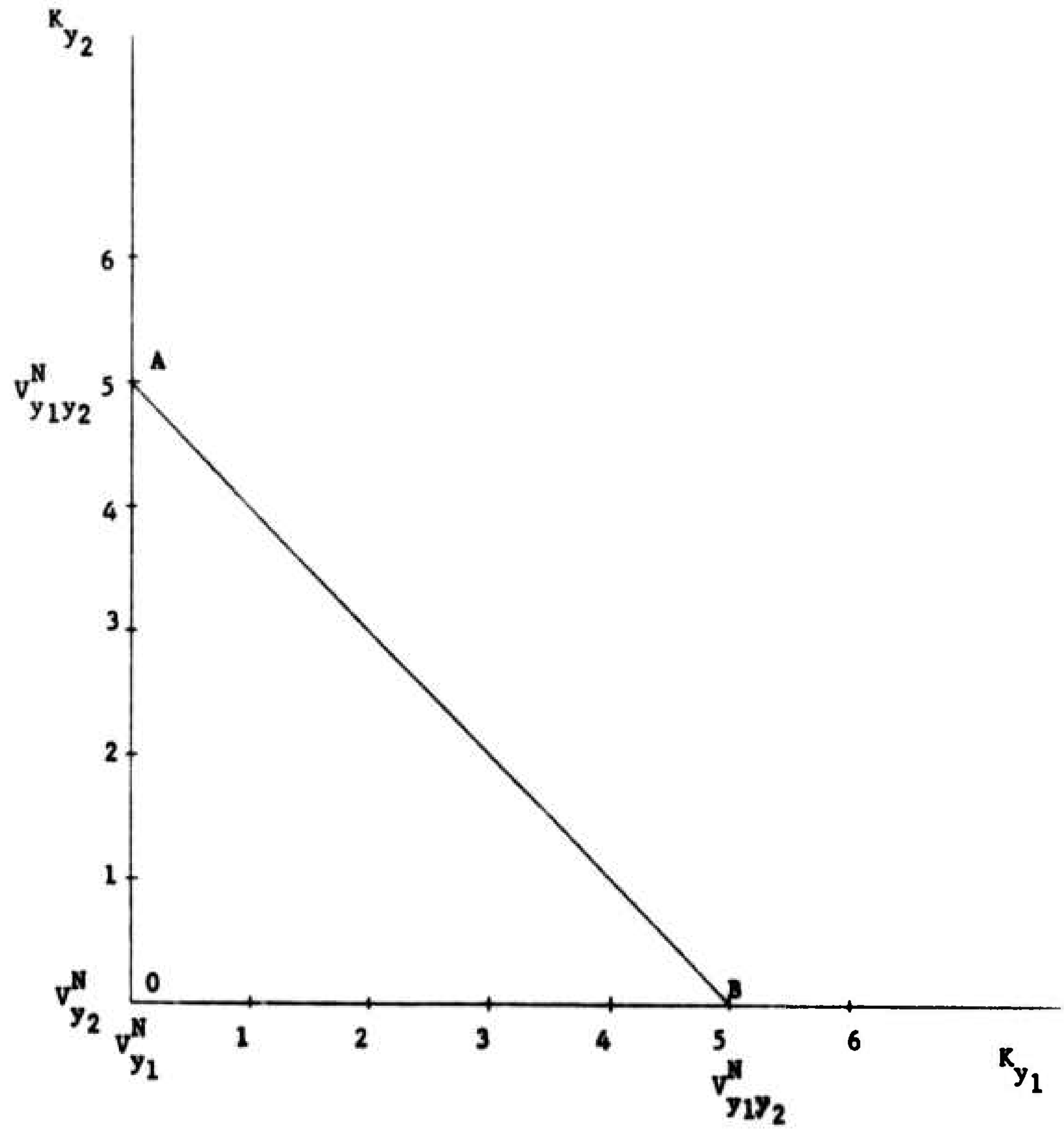


Figure 1.1. Price diagram for decision problem one

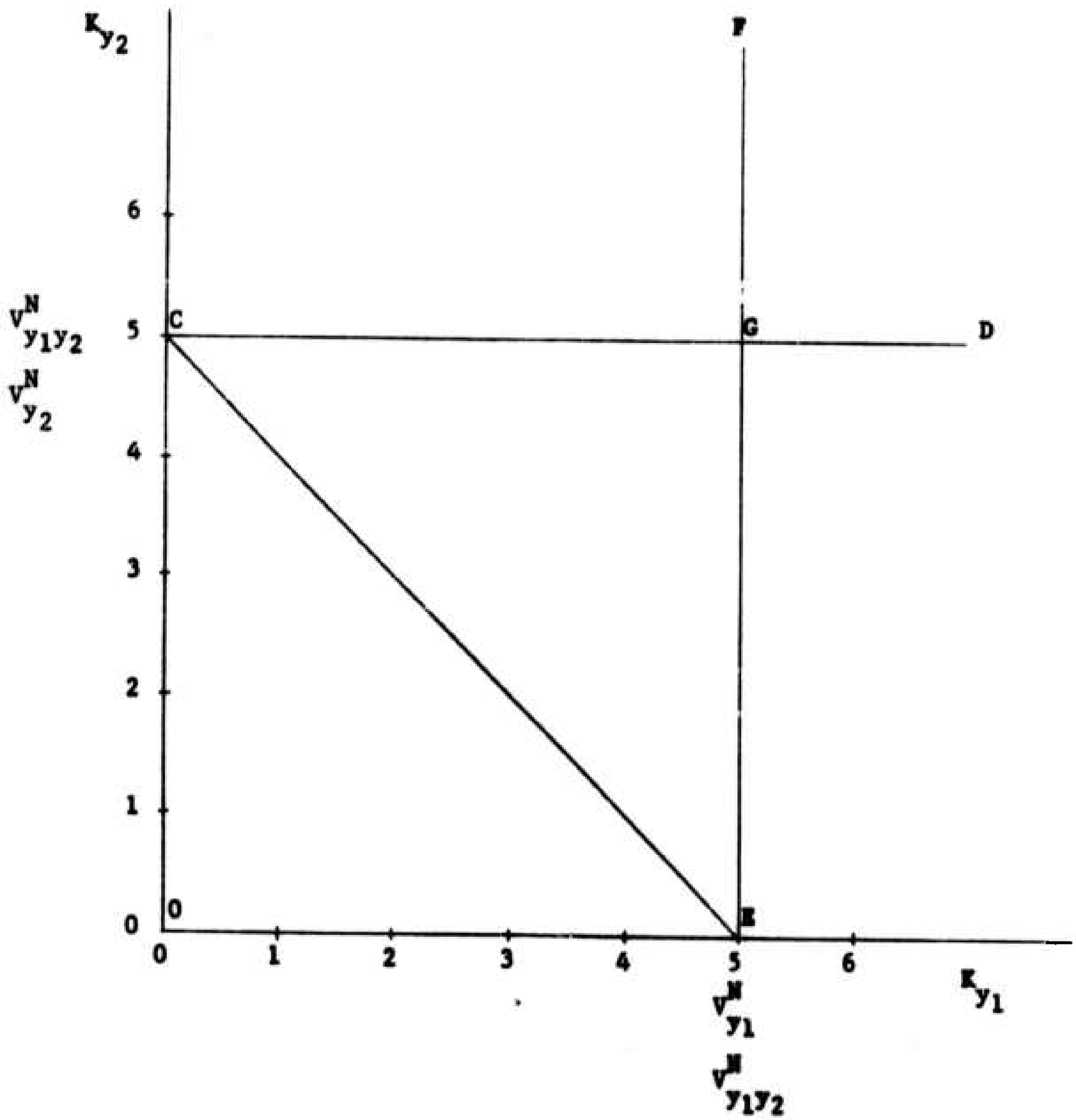


Figure 1.2. Price diagram for decision problem two

corresponding to y_1 when $K_{y_1} \leq V^N_{y_1}$, the information corresponding to y_2 when $K_{y_2} \leq V^N_{y_2}$, and both pieces of information when $(K_{y_1} + K_{y_2}) \leq V^N_{y_1 y_2}$. (It is possible that the cost of learning two pieces of information does not equal the sum of the costs of learning each individually. The case of non-additive prices is discussed in Chapter 5. For the moment we will assume that the prices are additive.)

These decision rules are shown in Fig. 1.1 for the first decision problem and Fig. 1.2 for the second decision problem. In the first decision problem $V^N_{y_1} = V^N_{y_2} = 0$, so we will only accept one piece of information by itself when it is free. However, $V^N_{y_1 y_2} = 5$, so we are willing to buy both pieces of information simultaneously whenever their prices are represented by a point below and to the left of the line A-B in Fig. 1.1.

Similarly, in the second decision problem we are willing to pay K_{y_1} to learn y_1 for any pair of prices represented by a point to the left of the line E-F in Fig. 1.2. We are also willing to pay K_{y_2} to learn y_2 for any pair of prices represented by a point below the line C-D, and we are willing to pay $(K_{y_1} + K_{y_2})$ to learn both pieces of information for any pair of prices below and to the left of the line C-E. Actually we would do better to buy only one piece of information for any pair of prices represented by a point within the triangle O-C-E since the second piece of information cannot raise our expected profit, but the idea here is that buying both pieces of information is preferable to not buying any information for such prices.

The interesting thing about the diagram in Fig. 1.1 is that there are prices represented by points in the interior of the triangle O-A-B such that we would not be willing to buy either piece of information

individually even though we would be willing to buy both pieces of information simultaneously. Pairs of prices with this property exist whenever $V^N_{y_1y_2}$ exceeds the sum of $V^N_{y_1}$ and $V^N_{y_2}$.

On the other hand, in Fig. 1.2 there are pairs of prices represented by points in the interior of the triangle C-E-G such that we would be willing to buy either piece of information individually, but we would not be willing to buy both pieces of information simultaneously. Pairs of prices with this property exist whenever $V^N_{y_1y_2}$ is less than the sum of $V^N_{y_1}$ and $V^N_{y_2}$.

The two decision problems have the same, simple probability structure. The observables y_1 and y_2 are independent random variables with the same probability mass functions in both problems. The only thing that is different in the two problems is the profit function. This is a demonstration of the fact that the relative values of $V^N_{y_1}$, $V^N_{y_2}$, and $V^N_{y_1y_2}$ are not completely determined by the probability structure of the problem. Many of the important linkages between various observables, that determine the relative value of each observable, occur through the profit function. As we shall see in Chapter 2, it is possible for the random variables representing two pieces of information to be independent, and yet the value of learning one observable can depend on the other observable. This dependence is the phenomenon that distinguishes sequential information from individual and simultaneous information.

When we have the option of learning information sequentially, we can spread our decisions out in time and evaluate the consequences of one decision before going on to make another. Although sequential decisions are inherently spread out in time, we will assume that the span of time is short enough to allow us to ignore the effects of time preference, such

as discounting. Since the content of one piece of information can affect the value of another, our decision to learn the second observable can vary with what we learn from the first. Furthermore, since learning one observable can cause us to change our decision about learning another, the first observable has a value above and beyond the value that we would determine by considering only individual and simultaneous purchases. Any time a piece of information can affect our subsequent decisions it is valuable. However, if we buy all of our information at once--either an individual purchase or the simultaneous purchase of several pieces of information--we have no subsequent decisions to make.

In terms of price diagrams, such as those in Figs. 1.1 and 1.2, we shall see that there are additional decision boundaries corresponding to the decision rules for buying information sequentially. These decision rules are not as simple as those for individual and simultaneous information, and the decision boundaries in the price diagram can be described by complicated, nonlinear equations. The decision boundaries for sequential information will usually allow us to purchase information at sets of prices that would not be advantageous if we considered only individual and simultaneous purchases of information. Graphically this means that the decision boundaries for sequential information can lie above and to the right of the corresponding boundaries for individual and sequential information, such as those shown in Figs. 1.1 and 1.2. This fact makes the question of whether or not $V^N_{y_1 y_2}$ exceeds the sum of $V^N_{y_1}$ and $V^N_{y_2}$ irrelevant. In most decision problems the decision about how much information to purchase will be determined by the value of sequential information, and not by the values of individual or simultaneous information. This idea is explored in some detail in the following chapters.

Different Types of Sequential Information Problems

It is possible to divide sequential information problems into a number of categories. Each category will be discussed in one of the following chapters.

When the two simple decision problems were discussed above, it was assumed that the cost of learning two observables was equal to the sum of the costs of learning each observable separately. When this assumption is valid for all combinations of information purchases, we say that the prices are additive. However, there are many decision problems for which the prices are not additive. For example, if we set up a research program to obtain one piece of information, we may find that the same facilities and expertise can be used to learn something else. In this case we can obtain both pieces of information for less than the sum of the costs of learning each one separately. Consequently, we can divide all sequential decision problems into those with additive and non-additive prices.

Another distinction based on the cost of information is whether or not we know the prices with certainty. If we do not know exactly what it will cost to learn a given observable, we can assess our uncertainty about the price in the form of a probability density function. For example, we might want to have a mechanic test our automobile to see if anything is wrong with it. However, before the mechanic actually does the work, we may be uncertain about how long it will take him to complete the test, and thus we will not know how much he will charge us for his labor. In a case like this, we will have to incorporate our uncertainty about the price into the determination of the value of each piece of information.

We can also distinguish between what is called perfect and imperfect

information. Perfect information tells us the exact value of one or more of the arguments of the profit function. In other words, if our profit depends on the number of items sold, the profit per item, and the manufacturing cost, being told one of these quantities would constitute perfect information about that quantity. On the other hand, conducting a market survey would give us imperfect information since it would change our state of information about the level of sales, but it would not tell us exactly how many items will be sold. In some ways the distinction between perfect and imperfect information is artificial because all information can be regarded as perfect information about something. The market survey tells us exactly what percentage of the group interviewed would be willing to buy our product. We shall see in Chapter 3 that the same problem formulation can be used for perfect and imperfect information. However, problems that deal only with perfect information can be formulated in a simpler manner, and they are easier to understand. For this reason the two types of information are discussed separately.

The following chapter deals with the simplest type of sequential information problem: the case of additive, certain prices for perfect information. A specific decision problem, in which we must bid for a contract, is used to explain the effect of sequential information purchasing and to show the procedure used to determine the value of sequential information. Much of the notation that is used in the rest of the work is developed in Chapter 2.

In Chapter 3 a problem in weather forecasting is used to illustrate the case of additive, certain prices for for imperfect information. The close relationship between problems with perfect and imperfect information

is discussed at the end of Chapter 3. Chapter 4 generalizes the procedures used in Chapters 2 and 3 for additive, certain prices. Several proofs are presented in Chapter 4 to show that the properties of the value of sequential information illustrated in Chapters 2 and 3 apply to all sequential information problems.

Chapter 5 deals with the case of uncertain prices by extending the bidding example introduced in Chapter 2. We show that under certain conditions we can analyze a problem with additive, uncertain prices by assuming that the prices are equal to their expected values. The question of non-additive prices is also discussed in Chapter 5.

CHAPTER 2

SEQUENTIAL PERFECT INFORMATION WITH ADDITIVE, CERTAIN PRICES: A BIDDING PROBLEM REVISITED

The concepts involved in the value of sequential perfect information can be best introduced with a simple example. The example chosen is one discussed in some detail by R. A. Howard [2 , 3]. It addresses the problem of how to choose the best bid when faced with uncertainty about your own production costs and your competitors' lowest bid. In this problem the profit from a contract secured by competitive bidding depends on two independent, continuous random variables: production cost and lowest competitive bid. The conclusions derived from the following calculations apply whether or not the random variables are independent, but the results are less intuitive and more striking in the case where they are independent. It is easy to show that the procedures developed in this example can be applied to discrete, as well as continuous, random variables. It can also be shown that the procedures and results apply when there are more than two random variables, but, in that case, the calculations become quite long and difficult. The bidding example does not consider utility functions or risk preference but the extension of the following calculations to include these concepts is straightforward. The prices of the observables are assumed to be additive and certain.

Notation

Although the bidding problem has been stated and solved by Howard, it will be restated briefly here so that this work is self-contained.

Restating the problem also makes it possible to adopt a notation slightly different from the one which Howard uses. Although Howard's inferential notation is very useful and has broad applications, the following notation, based on operators, is somewhat clearer and more compact when dealing with the value of information.

The inferential notation used by Howard shows explicitly the state of information of the person making the bidding decision. In this paper the decision maker's prior state of information will not be shown explicitly, and the state of his information at any time during the decision process will be shown by the use of certain operators. The notation uses the following definitions:

$\mathcal{S}$ = a state of information on which probability assignments are made,

$\mathcal{E}$ = the decision maker's prior state of information,

$\{x|\mathcal{S}\}$ = the density function (for continuous random variables) or mass function (for discrete random variables) of the random variable x, given the state of information $\mathcal{S}$,

$\langle x|\mathcal{S}\rangle = \int x\{x|\mathcal{S}\}\,dx$ = the expected value of x given the state of information $\mathcal{S}$,

$\bar{x} = \langle x|\mathcal{E}\rangle$ = the expected value of x given the decision maker's prior state of information,

$\pi(x_1, \ldots, x_n, c)$ = a profit function that depends on the state and control variables,

c = a control or decision variable (possibly vector-valued) upon which the profit depends,

x_i = a state variable; a random variable upon which the profit depends. (If we are dealing with perfect information, x_i is also an observable.)

y_j = an observable; a random variable whose actual value we have an opportunity to learn ,

$\underset{x_i}{E}\,\pi(x_1, \ldots, x_n, c) = \int \pi(x_1, \ldots, x_n, c) \left\{x_i \middle| \mathcal{E}\right\} dx$ = the expected value of the profit with respect to x_i ,

$\max_c \pi(x_1, \ldots, x_n, c)$ = the maximum of the profit with respect to c ,

K_{x_i} = cost of learning x_i ,

$V^N_{x_i}$ = the value of learning x_i given that no additional information is purchased; the value of individual information about x_i (the superscript N means "no additional information") ,

$V^N_{x_i x_j \ldots x_r}$ = the value of learning x_i , x_j , ..., and x_r simultaneously given that no additional information is purchased; the value of simultaneous information about x_i , x_j , ..., and x_r ,

$V^R_{x_i} = V^N_{x_1 \ldots x_n} - K_{x_1} - \ldots - K_{x_{i-1}} - K_{x_{i+1}} - \ldots - K_{x_n}$, the residual value of learning x_i when all of the available information is purchased simultaneously; the residual value of x_i , (the superscript R stands for "residual value") ,

V_{x_i} = the value of learning x_i when additional information can be purchased; the value of sequential information about x_i .

The costs and values above refer to perfect information since the quantities in the subscripts are state variables. If the subscripts are observables, the costs and values can refer to imperfect information. In this chapter we will be concerned with perfect information.

$\underset{x_i}{E}$ and $\underset{c}{\max}$ are the two basic operators that are used in the following discussion. By themselves, these operators do not show the state of information of the decision maker at the time that the expected or maximum value is determined. However, when a string of these operators is applied to $\pi(x_1, \ldots, x_n, c)$ such that there is one expectation operator for each random variable and one maximization operator for the control variable, the string of operators contains all of the information necessary to show the decision maker's state of information.

For example, the following expression represents the expected profit if the decision maker chooses the best setting of the control variable without any information about the state variables other than his prior state of information

$$\max_c \underset{x_1}{E} \underset{x_2}{E} \cdots \underset{x_n}{E} \pi(x_1, \ldots, x_n, c)$$

In inferential notation this would be written

$$\max_c(<<\ldots<\pi(x_1, \ldots, x_n, c)|x_1, \ldots, x_{n-1}, c, \varepsilon>|\ldots>|c, \mathcal{E}>)$$

As the inferential notation shows, each expectation is conditioned on a number of state and control variables. In the operator notation this means that $\underset{x_i}{E}$ is interpreted as the expected value with respect to x_i given all of the control or random variables that appear as subscripts of other operators that operate on $\underset{x_i}{E}$. We could show this dependency in

the operator subscripts by writing

$$\max_c \underset{x_1}{E} \cdots \underset{x_n}{E} \pi(x_1, \ldots, x_n, c) = \max_c \underset{x_1|c}{E} \cdots \underset{x_n|c,x_1,\ldots,x_{n-1}}{E} \pi$$

However, this is unnecessarily cumbersome.

When we are dealing with perfect information and independent state variables, there is no need to worry about dependencies among the state and control variables. However, in many problems the state variables are dependent. Furthermore, we must consider dependent random variables when we look at imperfect information. We can still use the operator notation as long as we are careful to interpret it correctly.

At times we will want to write expressions in operator notation that do not include all of the control or random variables as subscripts of an operator. When we are dealing with dependent random variables, we need to know whether or not the expectations in an expression are conditioned on the random variables that are not present. For example, $\max_c \underset{x_2}{E} \pi(x_1,x_2,c)$ could mean

$$\max_c\left(\int_{-\infty}^{\infty} \pi(x_1,x_2,c) \ \{x_2|x_1,c,\mathcal{E}\} \ dx_2\right)$$

or

$$\max_c\left(\int_{-\infty}^{\infty} \pi(x_1,x_2,c) \ \{x_2|c,\mathcal{E}\} \ dx_2\right)$$

The question is whether to use the conditional or marginal density function when determining the expected value of π with respect to x_2. In keeping with the rule that $\underset{x_i}{E}$ is interpreted as the expected value with respect to x_i given all of the control or random variables that appear as subscripts of other operators that operate on $\underset{x_2}{E}$, we define

$$\max_{c} \mathop{E}_{x_2} \pi(x_1,x_2,c) = \max_{c} \left(\int_{-\infty}^{\infty} \pi(x_1,x_2,c) \{x_2 | c, \mathcal{E}\} \, dx_2 \right.$$

The expectation is not conditioned on x_1 because x_1 does not appear as the subscript of another operator that operates on E_{x_2} .

Suppose we had wanted the expectation in the preceding expression to be with respect to x_2 conditioned on x_1 . We have no way to express this in operator notation using only the maximization and expectation operators. Therefore we define a pseudo-operator, $\mathop{D}_{x_1}$, that does not do anything to its argument, but indicates that all subsequent expectations are conditioned on x_1 . Thus

$$\mathop{D}_{x_1} \max_{x} \mathop{E}_{x_2} \pi(x_1,x_2,c) = \max_{c} \left(\int_{-\infty}^{\infty} \pi(x_1,x_2,c) \{x_2 | x_1 c, \mathcal{E}\} \, dx_2 \right)$$

We will not need the dependency pseudo-operator $\mathop{D}_{x_1}$ in this chapter because we are dealing with perfect information and the state variables are independent. However, we will need this notation in the following chapters.

Now we can use the operator notation to determine the value of information. Assume that the decision maker fortuitously learns the true value of x_i , and does not intend to buy additional information. Then his expected profit is

$$\mathop{D}_{x_i} \max_{c} \mathop{E}_{x_1} \cdots \mathop{E}_{x_{i-1}} \mathop{E}_{x_{i+1}} \cdots \mathop{E}_{x_n} \pi(x_1,\ldots,x_n,c)$$

This expression is a function of x_i . Before the decision maker learns the actual value of x_i , his expected profit is

$$\mathop{E}_{x_i} \max_{c} \mathop{E}_{x_1} \cdots \mathop{E}_{x_{i-1}} \mathop{E}_{x_{i+1}} \cdots \mathop{E}_{x_n} \pi(x_1,\ldots,x_n,c)$$

This is simply the expected value with respect to x_i of the preceding expression. If the decision maker is trying to maximize his expected profit, he will view the value of perfect information about x_i as the increase in his expected profit. Thus the value of perfect information about x_i, given that no additional information will be received, is

$$V^N_{x_i} = (\underset{x_i}{E} \max_c \underset{x_1}{E} \cdots \underset{x_{i-1}}{E} \underset{x_{i+1}}{E} \cdots \underset{x_n}{E} - \max_c \underset{x_1}{E} \cdots \underset{x_n}{E})\pi$$

Similarly, if the decision maker knows that he will learn x_i and x_j --but no additional information--before he has to choose a value for c, his expected profit prior to receiving the information is

$$\underset{x_i}{E} \underset{x_j}{E} \max_c \underset{x_1}{E} \cdots \underset{x_{i-1}}{E} \underset{x_{i+1}}{E} \cdots \underset{x_{j-1}}{E} \underset{x_{j+1}}{E} \cdots \underset{x_n}{E} \pi$$

In this case the value of perfect information about these two state variables is

$$V^N_{x_i x_j} = (\underset{x_i}{E} \underset{x_j}{E} \max_c \underset{x_1}{E} \cdots \underset{x_{i-1}}{E} \underset{x_{i+1}}{E} \cdots \underset{x_{j-1}}{E} \underset{x_{j+1}}{E} \cdots \underset{x_n}{E}$$

$$- \max_c \underset{x_1}{E} \cdots \underset{x_n}{E})\ \pi(x_1, \ldots, x_n, c)$$

The Bidding Problem

The operator notation will now be used to solve the bidding problem. Suppose our company must make a single sealed bid for a contract to produce a specific quantity of some commodity. The company's objective in bidding is to maximize its expected profit. We know exactly what must be done to meet the terms of the contract but we are uncertain about how much it will cost. Let p be our production cost and let $\{p|\mathcal{E}\}$ represent our probability density function on p based on our prior state of

information. The techniques for assessing $\{p|\mathcal{E}\}$ are discussed in several papers [1,4,11]. We shall assume that the decision maker's prior state of information about p has already been determined, and that the resulting probability density function is a uniform distribution between zero and one.

Similarly, let ℓ be our competitors' lowest bid and let $\{\ell|\mathcal{E}\}$ be the probability density function on ℓ based on our prior state of information. Assume that $\{\ell|\mathcal{E}\}$ is a uniform distribution between zero and two, and that the decision maker has assessed p and ℓ as independent random variables. The prior probability density functions for p and ℓ are shown in Fig. 2.1. Although we do not write these distributions explicitly in our notation, they will be used in the calculations whenever we need to determine the expected value of some function.

The profit in this bidding situation obviously depends on ℓ and p. It also depends on the amount we bid, b. In this problem ℓ and p are the state variables and b is the control variable (decision variable). If our bid is higher than the lowest competing bid, then we will not get the contract or incur any production costs, so our profit will be zero. If our bid is less than the lowest competing bid we will win the contract and receive an amount of money equal to b. However, we will pay a production cost to meet the terms of the contract, so our profit will be the difference between b and p. Thus the profit is given by

$$\pi(p,\ell,b) = \left\{\begin{matrix} b-p & : & b < \ell \\ \\ 0 & : & b \geq \ell \end{matrix}\right\}$$

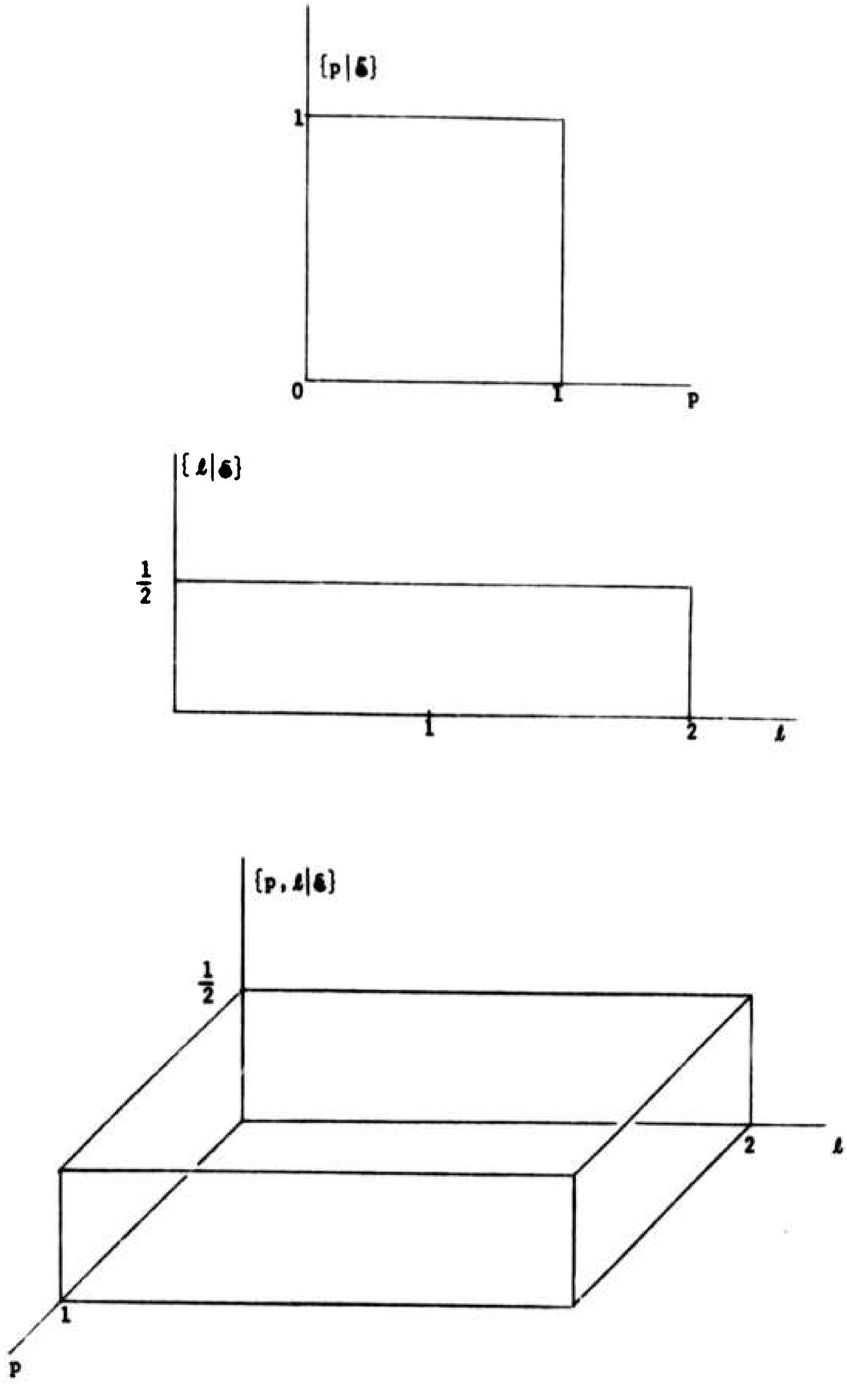


Figure 2.1. Probability density functions for production cost and lowest competing bid

It is possible to view the profit π as a random variable since it is a function of the random variables p and ℓ. It is also possible to consider the value of information as a random variable. We can derive probability density functions for the profit and the value of information, as shown in Appendix A. However, that will not be necessary for this example.

If we do not receive any information about p or ℓ, then the expected profit is

$$\max_b \mathop{E}_p \mathop{E}_\ell \pi(p,\ell,b) = \max_b \mathop{E}_p \mathop{E}_\ell \begin{Bmatrix} b-p : b < \ell \\ \\ 0 : b > \ell \end{Bmatrix} = 27/96$$

The maximum profit occurs when we make a bid of 5/4 .

The Values of Individual and Simultaneous Information

If we are given perfect information about our production cost, and do not intend to buy information about the lowest competing bid, then our expected profit prior to learning this information is

$$\mathop{E}_p \max_b \mathop{E}_\ell \pi(p,\ell,b) = \mathop{E}_p \max_b \mathop{E}_\ell \begin{Bmatrix} b-p : b < \ell \\ \\ 0 : b \geq \ell \end{Bmatrix} = 28/96$$

The maximum profit occurs when we make a bid equal to $(1 + p/2)$. Using the result above for the expected profit we find that the value of perfect individual information about p is

$$V_p^N = (\mathop{E}_p \max_b \mathop{E}_\ell - \max_b \mathop{E}_p \mathop{E}_\ell)\ \pi(p,\ell,b) = 28/96 - 27/96 = 1/96$$

If we are given perfect information about our competitors' lowest bid, and do not intend to buy information about our production cost, then

our expected profit prior to learning this information is

$$\underset{\ell}{E}\ \underset{b}{\max}\ \underset{p}{E}\ \pi(p,\ell,b) = \underset{\ell}{E}\ \underset{b}{\max}\ \underset{p}{E} \begin{Bmatrix} b-p : b < \ell \\ \\ 0 \quad : b \geq \ell \end{Bmatrix} = 54/96$$

The maximum profit occurs when

$$b = \begin{Bmatrix} \ell^- : \ell > 1/2 \\ \\ \infty \quad : \ell \leq 1/2 \end{Bmatrix}$$

In other words, we should just underbid the competition whenever we learn that their bid will be high enough to allow us to have a positive expected profit. Otherwise we should not bid. The value of perfect individual information about ℓ is

$$V_\ell^N = (\underset{\ell}{E}\ \underset{b}{\max}\ \underset{p}{E} - \underset{b}{\max}\ \underset{p}{E}\ \underset{\ell}{E})\ \pi(p,\ell,b) = 54/96 - 27/96 = 27/96$$

If we are given perfect information about both p and ℓ simulteneously, then our expected profit prior to learning this information is

$$\underset{p}{E}\ \underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) = \underset{p}{E}\ \underset{\ell}{E}\ \underset{b}{\max} \begin{Bmatrix} b-p : b < \ell \\ \\ 0 \quad : b \geq \ell \end{Bmatrix} = 56/96$$

The maximum expected profit occurs when

$$b = \begin{Bmatrix} \ell^- : \ell > p \\ \\ \infty \quad : \ell \leq p \end{Bmatrix}$$

In other words, we should just underbid the competition whenever we learn that their bid will exceed our production costs. Otherwise we should not bid. The value of perfect simultaneous information about both p and ℓ is

$$V^N_{p\ell} = (\underset{p}{E}\,\underset{\ell}{E}\,\underset{b}{\max} - \underset{b}{\max}\,\underset{p}{E}\,\underset{\ell}{E})\ \pi(p,\ell,b) = 56/96 - 27/96 = 29/96$$

Now we can compare $V^N_{p\ell}$ with V^N_p and V^N_ℓ. As the examples in the introduction demonstrate, the value of simultaneous information about p and ℓ need not equal the sum of the values of information about p and ℓ individually. The calculations above show that

$$V^N_p + V^N_\ell = 1/96 + 27/96 = 28/96 \neq V^N_{p\ell}$$

To gain a better understanding of this situation define the purchase prices of perfect information about p and ℓ to be K_p and K_ℓ, respectively. We are willing to buy perfect information about p when $K_p < V^N_p$, about ℓ when $K_\ell < V^N_\ell$, and about both p and ℓ when $(K_p + K_\ell) < V^N_{p\ell}$. (There may be situations where the cost of learning perfect information about two random variables does not equal the sum of the prices of learning the two individually. The case of non-additive prices will be discussed in Chapter 5.) These decision rules are shown graphically in Fig. 2.2. Any pair of prices is represented by a point on the K_p-K_ℓ diagram, which we call the "price diagram." We will buy perfect information about p for any pair of prices represented by a point below the line A-B. We will buy perfect information about ℓ for any pair of prices represented by a point to the left of the line C-D. We will buy simultaneous information about p and ℓ for any pair of prices represented by a point below and to the left of the line E-F.

The interesting thing about Fig. 2.2 is that there are pairs of prices, represented by points in the interior of the triangle G-H-I, such that we would be unwilling to pay for perfect information about either p or ℓ separately, but we would be willing to pay the same

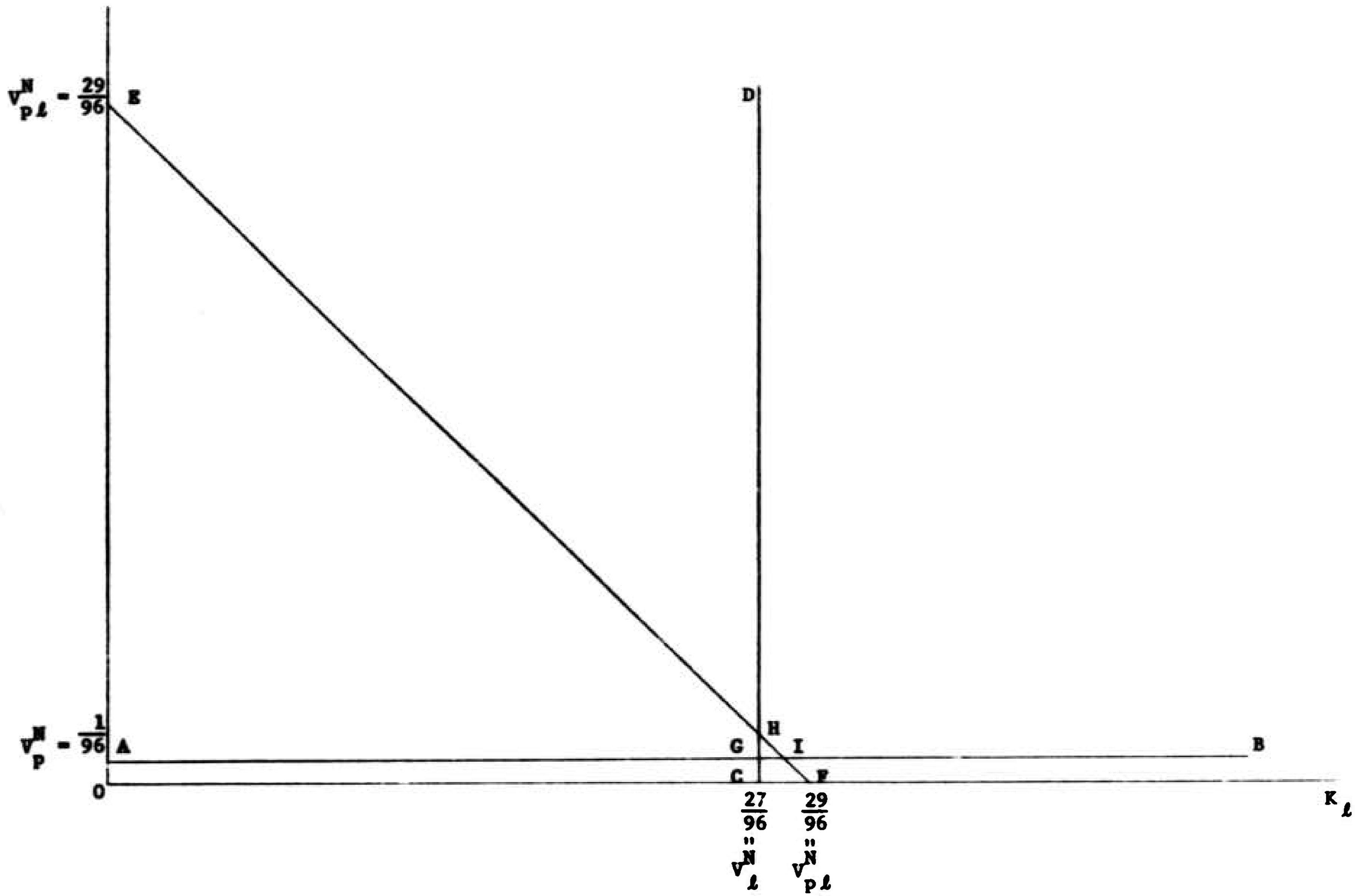


Figure 2.2. Price diagram for individual and simultaneous information

prices to learn the two simultaneously. This sort of diagram is useful in the following discussion where it will be shown that there are other pairs of prices, represented by points above and to the right of the boundary D-H-I-B, such that we will be willing to pay for perfect information about one or more of the random variables.

Although Fig. 2.2 shows that we would be willing to buy indivudal or simultaneous information about p or ℓ for any pair of prices represented by a point below and to the left of the boundary D-H-I-B, it does not tell us which information purchase is most advantageous. Assume for the moment that we can only buy pieces of information individually or simultaneously. We can determine which kind of information purchase is best by comparing the increase in expected profit in each case, since our objective is to maximize expected profit. The increase in expected profit is the difference between the value of the information and the cost of the information. Therefore we should pay for simultaneous information about p and ℓ when the following three conditions are met:

$$V^N_{p\ell} - K_p - K_\ell > V^N_p - K_p$$

$$V^N_{p\ell} - K_p - K_\ell > V^N_\ell - K_\ell$$

$$V^N_{p\ell} - K_p - K_\ell > 0$$

Simplifying these inequalities yields

$$K_\ell < V^N_{p\ell} - V^N_p = 29/96 - 1/96 = 28/96$$

$$K_p < V^N_{p\ell} - V^N_\ell = 29/96 - 27/96 = 2/96$$

$$K_p + K_\ell < V^N_{p\ell} = 29/96$$

Similar inequalities can be written when our best decision is to buy individual information about p or ℓ. These inequalities define the decision regions shown in Fig. 2.3.

Using Fig. 2.3 we can immediately determine our best response when we are offered perfect individual or simultaneous information about p and ℓ at prices K_p and K_ℓ respectively. As Fig. 2.3 shows, it is necessary to know the prices of both pieces of information to decide which observable to buy, even when only individual and simultaneous purchases are possible. We shall see that it is possible to draw a similar diagram when we have the option to buy information sequentially. When sequential information is available the decision regions become more complicated than those in Fig. 2.3, and the best information purchase at any pair of prices can be different from what it is in Fig. 2.3.

We can use $V^N_{p\ell}$, the value of simultaneous information about p and ℓ, to define two related quantities: $V^R_p(K_\ell)$ and $V^R_\ell(K_p)$. $V^R_p(K_\ell)$ is the residual value of learning p when we must buy information about p and ℓ simultaneously, and the information about ℓ costs K_ℓ. We call $V^R_p(K_\ell)$ the residual value of p. The maximum amount we will pay for this information is $(V^N_{p\ell} - K_\ell)$, and, by definition, the maximum amount that we will pay for information is the value of that information. Therefore we define

$$V^R_p(K_\ell) = V^N_{p\ell} - K_\ell$$

The superscript R stands for "residual value."

We are willing to buy information about p and ℓ simultaneously whenever

$$K_p \le V^R_p(K_\ell)$$

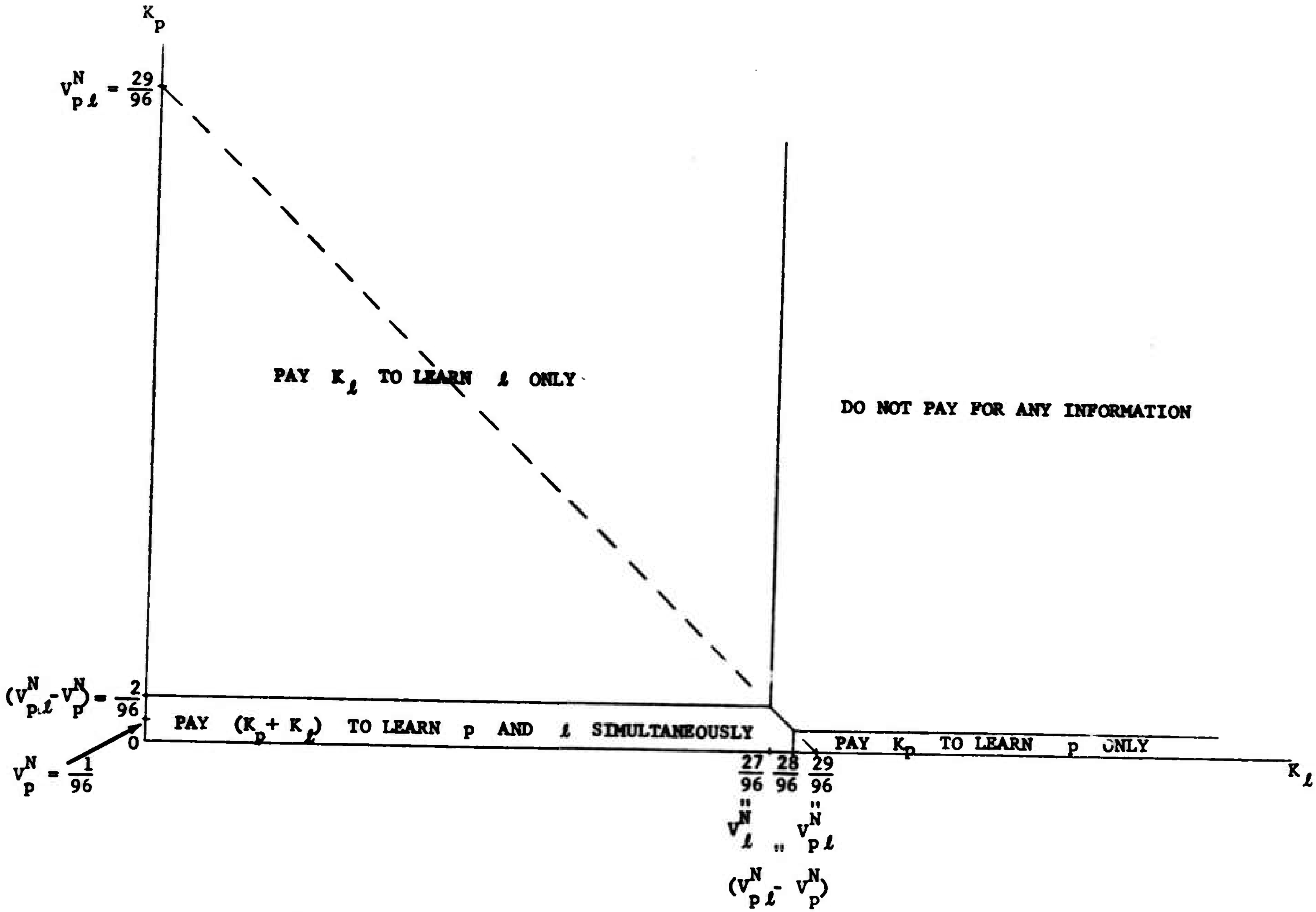


Figure 2.3. Decision regions for individual and simultaneous information

This is just another way to write the decision rule for buying simultaneous information about p and ℓ . Substituting in this inequality brings us back to the decision rule we found previously

$$K_p + K_\ell \leq V^N_{p\ell}$$

In exactly the same way we define

$$V^R_\ell(K_p) = V^N_{p\ell} - K_p$$

We should buy information about p and ℓ simultaneously whenever

$$K_\ell \leq V^R_\ell(K_p)$$

For the bidding example,

$$V^R_p(K_\ell) = 29/96 - K_\ell$$

$$V^R_\ell(K_p) = 29/96 - K_p$$

When we write $V^R_p(K_\ell)$ or $V^R_\ell(K_p)$ we are assigning all of the value of learning p and ℓ simultaneously to one of the observables. While this may seem artificial, it is useful in the discussion that follows. We shall see that the value of sequential information about p approaches V^N_p when K_ℓ is very large, and it approaches $V^R_p(K_\ell)$ when K_ℓ is small. Similarly, $V_\ell(K_p)$ approaches V^N_ℓ when K_p is very large and $V^R_\ell(K_p)$ when K_p is small.

We now want to consider the process of buying perfect information sequentially. We shall see that gaining perfect information about one state variable can tell us something about whether or not to pay for

perfect information about a second state variable, even though we have assumed that they are independent random variables.

The Value of Sequential Information about p

We shall now determine the value of perfect information about p when we know that subsequently we can pay K_ℓ for perfect information about ℓ. Assume for a moment that we already know the actual value of our production cost, and that we are trying to decide whether or not to buy perfect information about our competitors' lowest bid. Since p and ℓ are independent random variables, our probability density function for ℓ is unchanged by our knowledge of p. How much is the additional information about ℓ worth to us?

If we decide not to learn ℓ, then the expected profit as a function of p (which we already know) is

$$\max_b \mathop{E}_\ell \pi(p,\ell,b) = (1 - p/2)^2/2$$

If p and ℓ were dependent random variables we would have to write

$$\mathop{D}_p \max_b \mathop{E}_\ell \pi(p,\ell,b)$$

since the expectation is conditioned on the value of p that we learned previously. However, we can drop the $\mathop{D}_p$ pseudo-operator since p and ℓ are independent. The maximum expected profit occurs for a bid of

$$b = 1 + p/2$$

This expected profit is shown in Fig. 2.4a.

On the other hand, if we decide to learn ℓ, then the expected profit as a function of p is

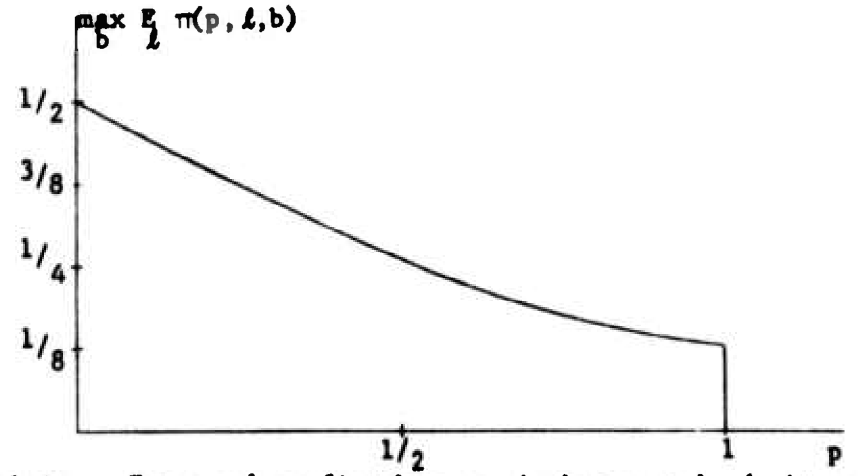


Figure 2.4a. Expected profit when p is known and ℓ is not learned

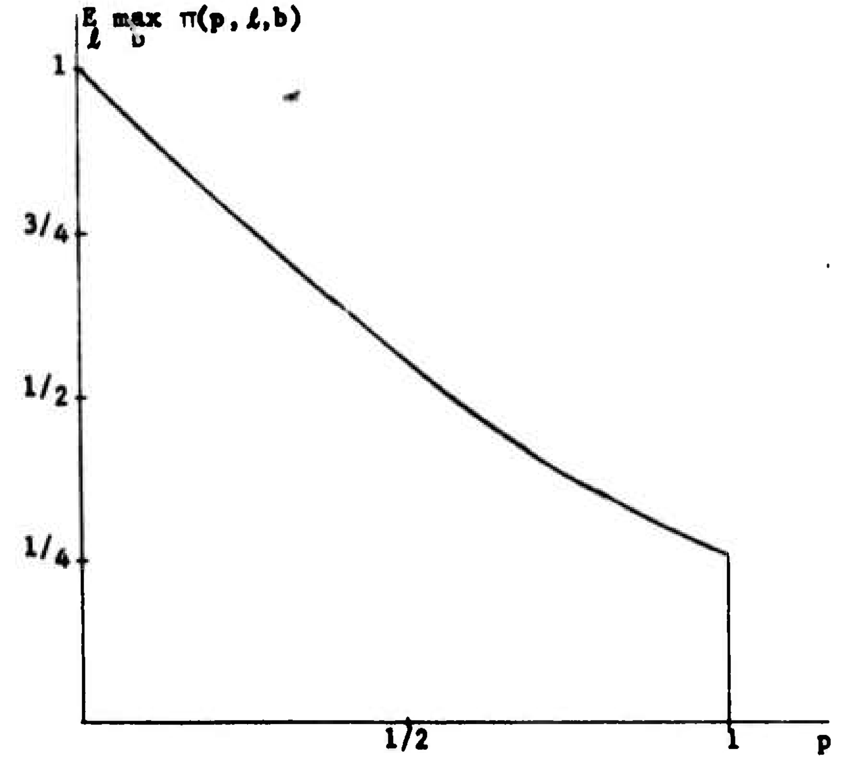


Figure 2.4b. Expected profit when p is known and ℓ is learned

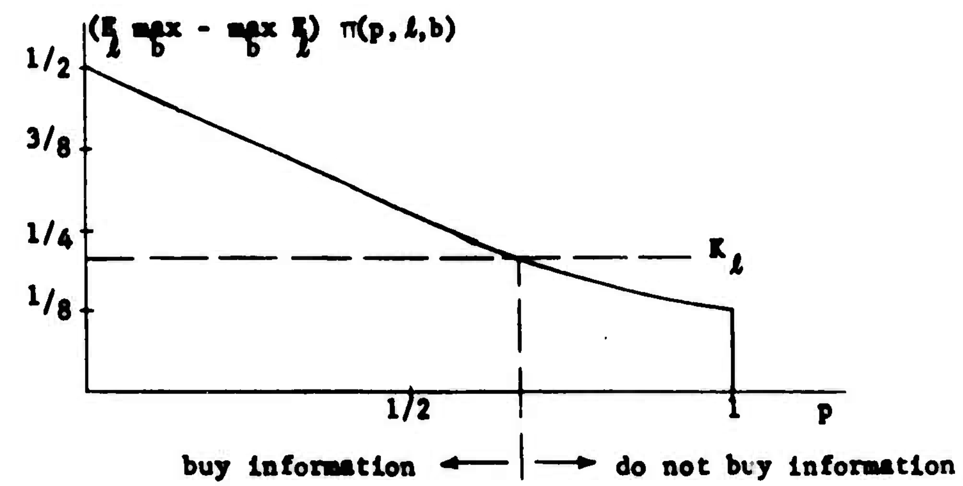


Figure 2.4c. Increase in expected profit caused by learning ℓ when p is known.

$$\underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) = (1 - p/2)^2$$

The maximum expected profit occurs for a bid of

$$b = \begin{Bmatrix} \ell^- & : \ell > p \\ \\ \infty & : \ell \leq p \end{Bmatrix}$$

This is shown in Fig. 2.4b. The increase in the expected profit caused by learning ℓ when we already know p is the difference between these two expressions:

$$(\underset{\ell}{E}\ \underset{b}{\max} - \underset{b}{\max}\ \underset{\ell}{E})\ \pi(p,\ell,b) = (1 - p/2)^2/2$$

Now suppose that the price of perfect information about ℓ is K_ℓ. Obviously we should pay this price when the increase in the expected profit exceeds K_ℓ; otherwise we should not buy the information. Thus we should pay K_ℓ for perfect information about ℓ when

$$K_\ell < (\underset{\ell}{E}\ \underset{b}{\max} - \underset{b}{\max}\ \underset{\ell}{E})\ \pi(p,\ell,b)$$

This decision rule is shown graphically for the bidding problem in Fig. 2.4c. It is obvious from this inequality, or from Fig. 2.4c, that for certain values of K_ℓ the decision to buy or not buy perfect information about ℓ depends on the value of p that we learned earlier. Actually there are three cases:

(1) If $K_\ell \leq 1/8$, we will always pay K_ℓ to learn ℓ, regardless of the value of p, because the increase in the expected profit will always exceed K_ℓ. In this case our expected profit as a function of p is

$$\mathop{E}_{\ell} \max_{b} \pi(p,\ell,b) - K_{\ell} = (1 - p/2)^2 - K_{\ell}$$

(2) If $1/8 < K_{\ell} < 1/2$, we may or may not pay K_{ℓ} to learn ℓ depending on the value of p . We will pay to learn ℓ when

$$K_{\ell} < (1 - p/2)^2/2 \quad \text{or} \quad 0 \leq p < 2(1 - \sqrt{2K_{\ell}})$$

Otherwise we will not pay to learn ℓ . Thus the expected profit as a function of p is

$$\max \left\{ \begin{array}{l} \mathop{E}_{\ell} \max_{b} \pi(p,\ell,b) - K_{\ell} \\ \\ \max_{b} \mathop{E}_{\ell} \pi(p,\ell,b) \end{array} \right\}$$

$$= \left\{ \begin{array}{ll} \mathop{E}_{\ell} \max_{b} \pi(p,\ell,b) - K_{\ell} & : 0 \leq p < 2(1 - \sqrt{2K_{\ell}}) \\ \\ \max_{b} \mathop{E}_{\ell} \pi(p,\ell,b) & : 2(1 - \sqrt{2K_{\ell}}) \leq p \leq 1 \end{array} \right\}$$

$$= \left\{ \begin{array}{ll} (1 - p/2)^2 - K_{\ell} & : 0 \leq p < 2(1 - \sqrt{2K_{\ell}}) \\ \\ (1 - p/2)^2/2 & : 2(1 - \sqrt{2K_{\ell}}) \leq p \leq 1 \end{array} \right\}$$

(3) If $K_{\ell} \geq 1/2$, we will never pay K_{ℓ} to learn ℓ , regardless of the value of p , because K_{ℓ} will always exceed the increase in the expected profit. In this case our expected profit as a function of p is

$$\max_{b} \mathop{E}_{\ell} \pi(p,\ell,b) = (1 - p/2)^2/2$$

These expressions give us the expected profit when we already know p . We can determine our expected profit before we learn p by taking

the expected value with respect to p of the previous results. Thus our expected profit when we know that we will receive perfect information about p, but before we actually receive the information, is

$$\mathop{E}_{p} \max \left\{ \begin{array}{l} \mathop{E}_{\ell} \max_{b} \pi(p,\ell,b) - K_{\ell} \\ \\ \max_{b} \mathop{E}_{\ell} \pi(p,\ell,b) \end{array} \right\}$$

$$= \left\{ \begin{array}{l} \mathop{E}_{p} (1 - p/2)^2 - K_{\ell} : K_{\ell} \leq 1/8 \\ \int_{0}^{2(1-\sqrt{2K_{\ell}})} [(1-p/2)^2 - K_{\ell}] \{p|\delta\}\, dp \\ \qquad + \int_{2(1-\sqrt{2K_{\ell}})}^{1} [(1-p/2)^2/2] \{p|\delta\}\, dp : 1/8 < K_{\ell} < 1/2 \\ \mathop{E}_{p} (1 - p/2)^2/2 : 1/2 \leq K_{\ell} \end{array} \right\}$$

$$= \left\{ \begin{array}{l} 56/96 - K_{\ell} : K_{\ell} \leq 1/8 \\ 60/96 + (4/3)K_{\ell}\sqrt{2K_{\ell}} - 2K_{\ell} : 1/8 < K_{\ell} < 1/2 \\ 28/96 : 1/2 \leq K_{\ell} \end{array} \right\}$$

Since our expected profit without any information is $\max_{b} \mathop{E}_{p} \mathop{E}_{\ell} \pi(p,\ell,b)$, we must subtract this quantity from the result above to get the value of perfect sequential information about p, or $V_p(K_{\ell})$.

$$V_p(K_{\ell}) = \mathop{E}_{p} \max \left\{ \begin{array}{l} \mathop{E}_{\ell} \max_{b} \pi(p,\ell,b) - K_{\ell} \\ \\ \max_{b} \mathop{E}_{\ell} \pi(p,\ell,b) \end{array} \right\} - \max_{b} \mathop{E}_{p} \mathop{E}_{\ell} \pi(p,\ell,b)$$

$$= \left\{ \begin{array}{l} 29/96 - K_{\ell} : K_{\ell} \leq 1/8 \\ 33/96 + (4/3)K_{\ell}\sqrt{2K_{\ell}} - 2K_{\ell} : 1/8 < K_{\ell} < 1/2 \\ 1/96 : 1/2 \leq K_{\ell} \end{array} \right\}$$

Thus if we are offered perfect information about p for a price K_p, we will accept the offer when $K_p \leq V_p(K_\ell)$. However, V_p depends on K_ℓ, the cost of perfect information about ℓ. Therefore we must know both K_p and K_ℓ before we can decide whether or not to accept the offer

In terms of K_p and K_ℓ, we should pay to learn p whenever

$$K_p \leq \begin{cases} 29/96 - K_\ell : K_\ell \leq 1/8 \\ 33/96 + (4/3)K_\ell\sqrt{2K_\ell} - 2K_\ell \quad : 1/8 < K_\ell < 1/2 \\ 1/96 : 1/2 \leq K_\ell \end{cases}$$

This decision rule is shown graphically in Fig. 2.5. We are willing to pay K_p to learn p for any pair of prices represented by a point below the boundary A-B-C-D. Before discussing the significance of this diagram, we will carry out another calculation, similar to the one above, to determine the value of perfect sequential information about ℓ.

The Value of Sequential Information about ℓ

This time we will determine the value of perfect information about ℓ when we know that we can subsequently pay K_p for perfect information about p. Assume for a moment that we know the actual value of our competitors' lowest bid, and that we are trying to decide whether or not to buy perfect information about our production cost. How much is the additional information about p worth to us?

If we decide not to learn p, then the expected profit as a function of ℓ is

$$\max_b \mathop{E}_p \pi(p,\ell,b) = \begin{cases} 0 & : 0 \leq \ell \leq 1/2 \\ \ell - 1/2 & : 1/2 < \ell \leq 2 \end{cases}$$

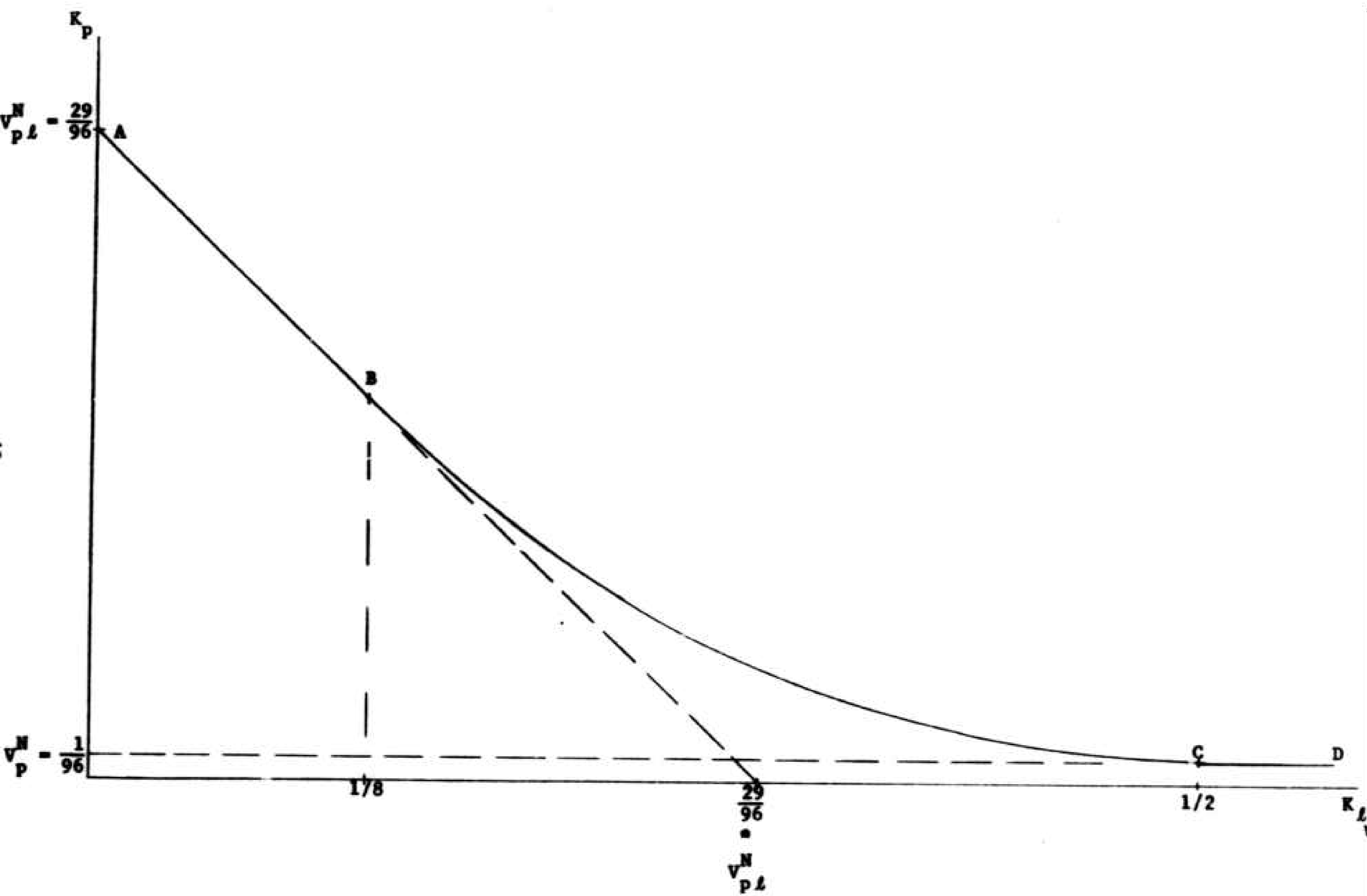


Figure 2.5. Value of sequential information about p as a function of K_ℓ

The maximum expected profit occurs for a bid of

$$b = \begin{cases} \ell^- & : \ell > 1/2 \\ \infty & : \ell \leq 1/2 \end{cases}$$

This expected profit is shown in Fig. 2.6a. On the other hand, if we are given perfect information about p , then the expected profit as a function of ℓ is

$$E \max \pi(p,\ell,b) = \begin{cases} \ell^2/2 & : 0 \leq \ell \leq 1 \\ \ell - 1/2 & : 1 < \ell \leq 2 \end{cases}$$

The maximum expected profit occurs for a bid of

$$b = \begin{cases} \ell^- & : \ell > p \\ \infty & : \ell \leq p \end{cases}$$

This is shown in Fig. 2.6b. The increase in expected profit caused by learning p when we already know ℓ is the difference between these two expressions:

$$(\underset{p}{E} \underset{b}{\max} - \underset{b}{\max} \underset{p}{E}) \pi(p,\ell,b) = \begin{cases} \ell^2/2 & : 0 \leq \ell \leq 1/2 \\ (1 - \ell)^2/2 & : 1/2 < \ell < 1 \\ 0 & : 1 \leq \ell \leq 2 \end{cases}$$

Now suppose that the price of perfect information about p is K_p . We should pay this price when the increase in the expected profit exceeds K_p ; otherwise we should not buy the information. Thus we should pay K_p for perfect information about p when

$$K_p < (\underset{p}{E} \underset{b}{\max} - \underset{b}{\max} \underset{p}{E}) \pi(p,\ell,b)$$

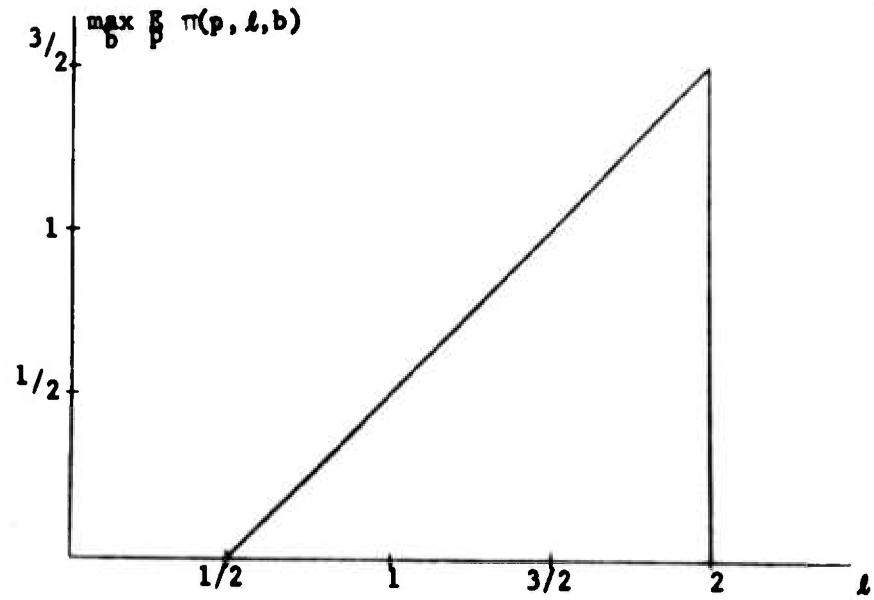


Figure 2.6a. Expected profit when ℓ is known and p is not learned

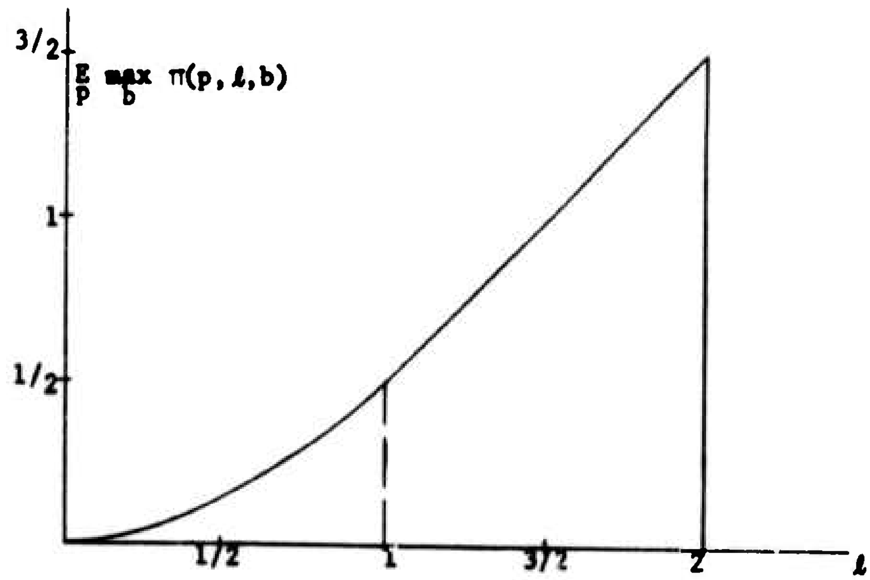


Figure 2.6b. Expected profit when ℓ is known and p is learned

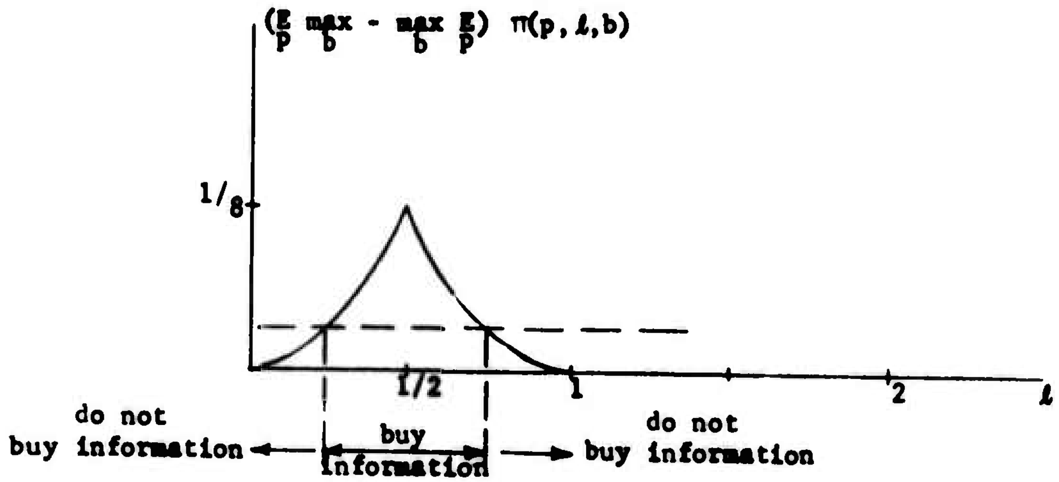


Figure 2.6c. Increase in expected profit caused by learning p when ℓ is known

This decision rule is shown graphically in Fig. 2.6c. It can be seen from this inequality, or from Fig. 2.6c, that for certain values of K_p the decision to buy or not buy perfect information about p depends on the value of ℓ that we learned earlier. This time there are only two cases:

(1) If $K_p < 1/8$, we may or may not pay K_p to learn p depending on the value of ℓ. We will pay to learn p when

$$K_p < \ell^2/2 \ , \quad \text{and} \quad K_p < (1 - \ell)^2/2$$

These inequalities can be rewritten as follows:

$$\sqrt{2K_p} < \ell < 1 - \sqrt{2K_p}$$

Otherwise we will not pay to learn p. Thus the expected profit, as a function of ℓ, is

$$\max\left\{\begin{matrix} \underset{p}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) - K_p \\ \underset{b}{\max}\ \underset{p}{E}\ \pi(p,\ell,b) \end{matrix}\right\}$$

$$= \left\{\begin{matrix} \underset{p}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) - K_p : \sqrt{2K_p} < \ell < 1 - \sqrt{2K_p} \\ \underset{b}{\max}\ \underset{p}{E}\ \pi(p,\ell,b) : 0 \le \ell \le \sqrt{2K_p}\ , \quad 1 - \sqrt{2K_p} \le \ell \le 2 \end{matrix}\right\}$$

$$= \left\{\begin{matrix} 0 & : 0 \le \ell \le \sqrt{2K_p} \\ \ell^2/2 - K_p & : \sqrt{2K_p} < \ell < 1 - \sqrt{2K_p} \\ \ell - 1/2 & : 1 - \sqrt{2K_p} \le \ell \le 2 \end{matrix}\right\}$$

(2) If $K_p \ge 1/8$, we will never pay K_p to learn p, regardless of the value of p, because K_p will always exceed the increase in the expected profit. In this case our expected

profit, as a function of ℓ , is :

$$\max E\,\pi(p,\ell,b) = \begin{cases} 0 & : 0 \leq \ell \leq 1/2 \\ \ell - 1/2 & : 1/2 < \ell \leq 2 \end{cases}$$

These expressions give us the expected profit when we already know ℓ . We can determine our expected profit before we learn ℓ by taking the expected value with respect to ℓ of the previous results. Thus our expected profit when we know that we will receive perfect information about ℓ , but before we actually get the information, is

$$\underset{\ell}{E}\max\begin{Bmatrix} \underset{p}{E}\,\underset{b}{\max}\,\pi(p,\ell,b) - K_p \\ \max E\,\pi(p,\ell,b) \end{Bmatrix}$$

$$= \begin{Bmatrix} \int_{\sqrt{2K_p}}^{1-\sqrt{2K_p}} (\ell^2/2 - K_p)\{\ell|\delta\}d\ell + \int_{1-\sqrt{2K_p}}^{2} (\ell-1/2)\{\ell|\delta\}d\ell : K_p < 1/8 \\ \int_{1/2}^{2} (\ell - 1/2)\,\{\ell|\delta\}d\ell : K_p \geq 1/8 \end{Bmatrix}$$

$$= \begin{Bmatrix} 56/96 + (2/3)K_p\sqrt{2K_p} - K_p/2 : K_p < 1/8 \\ 54/96 : K_p \geq 1/8 \end{Bmatrix}$$

Since the expected profit without any information is $\underset{b}{\max}\,\underset{p}{E}\,\underset{\ell}{E}\,\pi(p,\ell,b)$, we must subtract this quantity from the result above to get the value of perfect sequential information about ℓ , or $V_\ell(K_p)$.

$$V_\ell(K_p) = \underset{\ell}{E}\max\begin{Bmatrix} \underset{p}{E}\,\underset{b}{\max}\,\pi(p,\ell,b) - K_p \\ \underset{b}{\max}\,\underset{p}{E}\,\pi(p,\ell,b) \end{Bmatrix} - \max E\,E\,\pi(p,\ell,b)$$

$$= \begin{Bmatrix} 29/96 + (2/3)K_p\sqrt{2K_p} - K_p/2 : K_p < 1/8 \\ 27/96 : K_p \geq 1/8 \end{Bmatrix}$$

If we are offered perfect information about ℓ for a price K_ℓ, we should accept the offer when $K_\ell < V_\ell(K_p)$.

In terms of K_p and K_ℓ, we should pay K_ℓ to learn ℓ whenever

$$K_\ell \leq \begin{cases} 29/96 + (2/3)K_p\sqrt{2K_p} - K_p/2 : K_p < 1/8 \\ 27/96 : K_p \geq 1/8 \end{cases}$$

This decision rule is shown graphically in Fig. 2.7. We are willing to pay K_ℓ to learn ℓ for any pair of prices represented by a point to left of the boundary E-F-G.

Combining the diagrams in Figs. 2.2, 2.5, and 2.7 shows how these prices are interrelated. The combined diagram is shown in Fig. 2.8. From our previous results we know that we syould pay K_p to learn the actual value of p for any pair of prices represented by a point below the curve A-B-C-D. In addition we should pay K_ℓ to learn the actual value of ℓ for any pair of points to the left of the curve E-F-G. Therefore we should be willing to pay for perfect information about one of the state variables when offered any pair of prices represented by a point below or to the left of the boundary G-F-H-C-D. This set of prices, which we call the set of feasible prices, includes all of the pairs of prices that we found previously by considering the purchase of individual information about each of the state variables. It also includes all of the pairs of prices that we found by considering the simultaneous purchase of perfect information about both of the state variables.

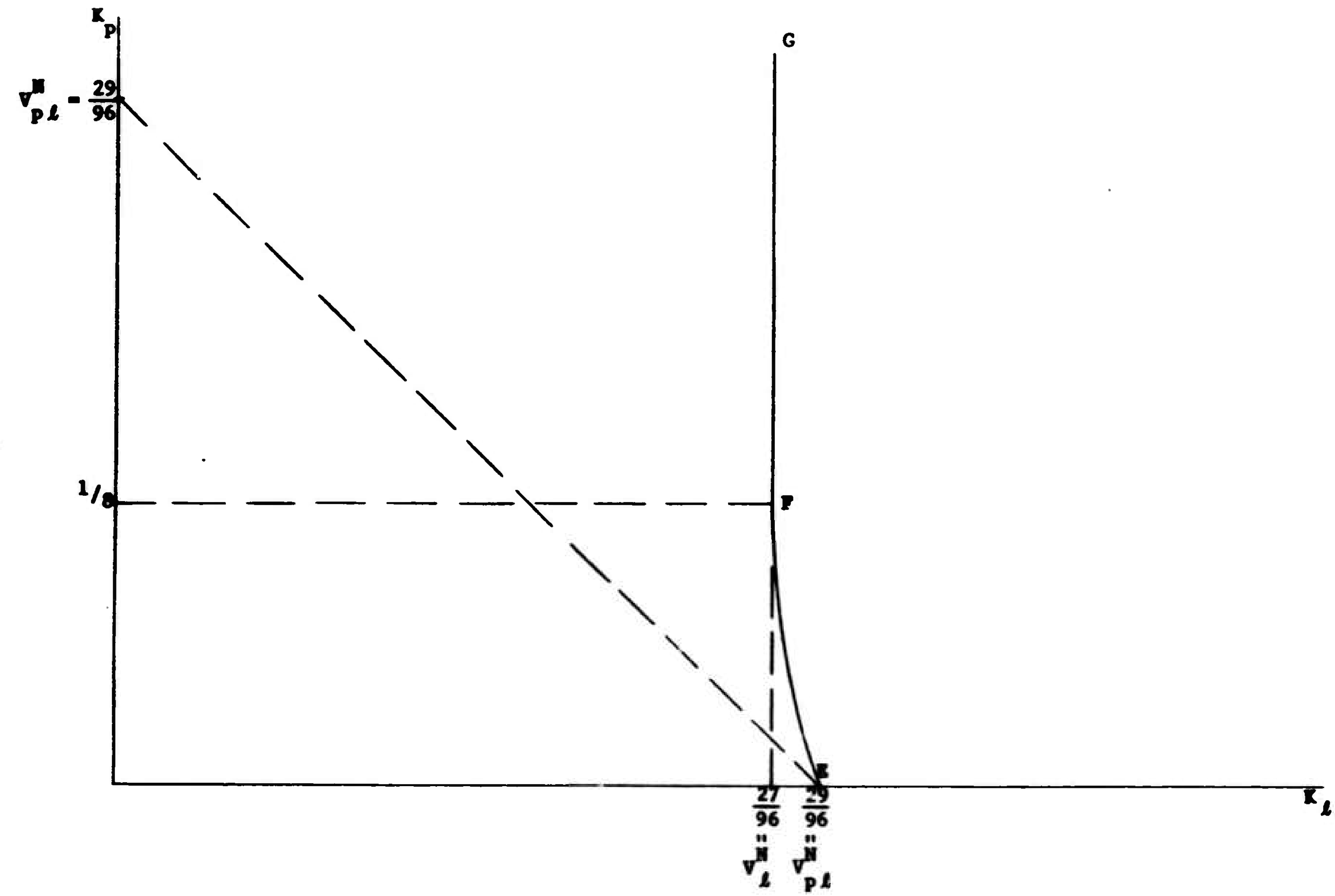


Figure 2.7. Value of sequential information about ℓ as a function of K_p

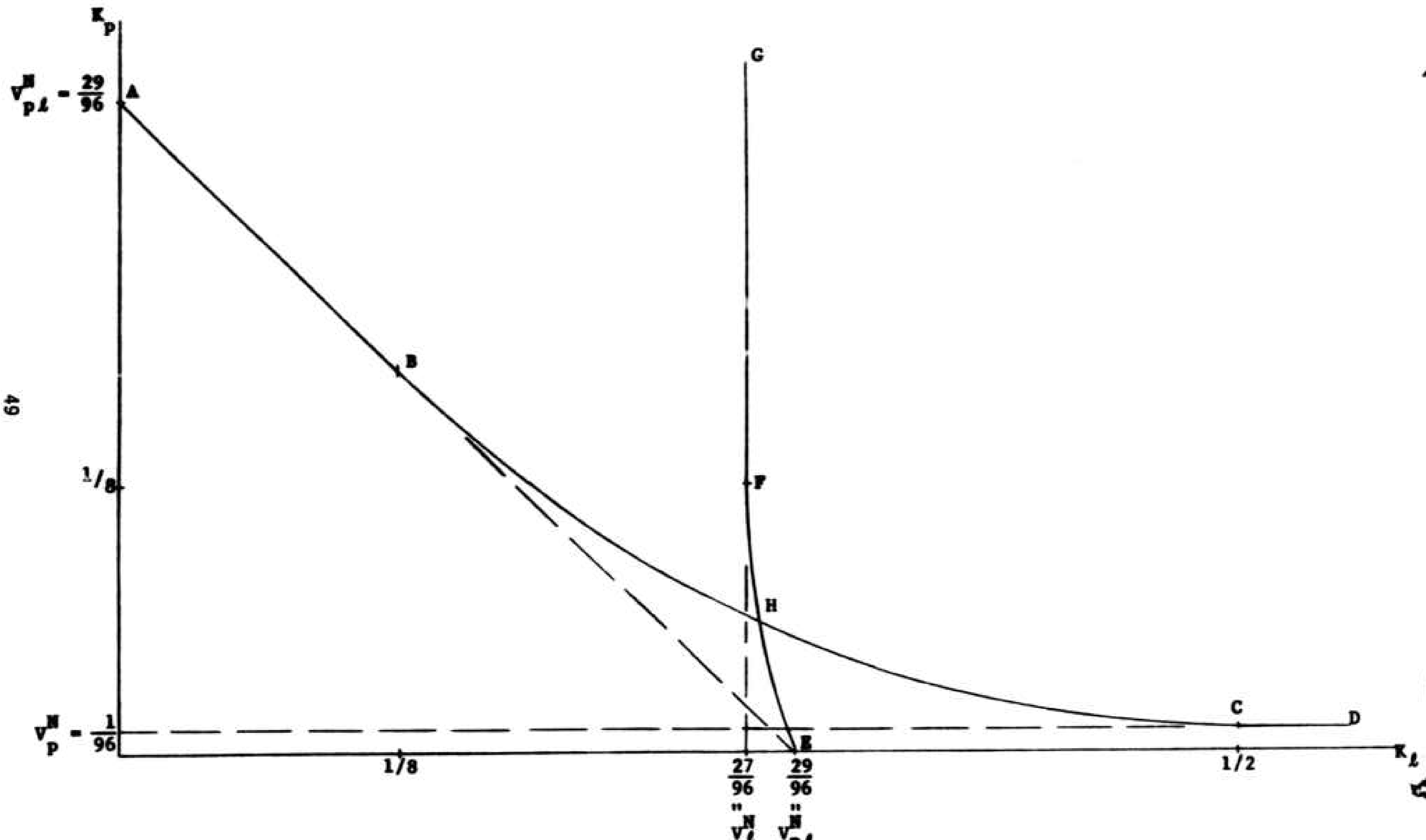


Figure 2.8. Price diagram for sequential information

The Relative Values of V_p^N , $V_p^R(K_\ell)$, and $V_p(K_\ell)$

Intuitively we would expect the set of feasible prices for sequential purchases of information to include the sets of feasible prices for individual and simultaneous purchases of information because learning observables individually and simultaneously are just special cases of learning the information sequentially. Purchasing information about a single state variable can be viewed as two sequential decisions: first to buy information about the state variable, and then not to buy any additional information. Similarly, purchasing information about both state variables can also be viewed as two sequential decisions: first to buy information about one state variable, and then to buy information about the other.

It is easy to see from the formulas derived previously that our intuition is correct. Consider the value of perfect information about our production costs. If we only buy perfect information about p , then the expected value of the information is

$$V_p^N = \underset{p}{E}[\max_b \underset{\ell}{E}\, \pi(p,\ell,b)] - \max_b \underset{p}{E}\, \underset{\ell}{E}\, \pi(p,\ell,b)$$

If we buy perfect information about p and ℓ simultaneously, then the residual value of learning p is

$$V_p^R(K_\ell) = \underset{p}{E}[\underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell] - \max_b \underset{p}{E}\, \underset{\ell}{E}\, \pi(p,\ell,b)$$

If we buy perfect information about p and ℓ sequentially, then the value of the information is

$$V_p(K_\ell) = \underset{p}{E} \max \left\{ \begin{array}{l} \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell \\ \max_b \underset{\ell}{E} \pi(p,\ell,b) \end{array} \right\} - \max_b \underset{p}{E} \underset{\ell}{E} \pi(p,\ell,b)$$

From the form of these equations it can be seen that $V_p(K_\ell)$ must be at least as large as V_p^N and $V_p^R(K_\ell)$ for any value of K_ℓ. The last term in all three expressions is the same, and the first term is the expected value with respect to p of some quantity. Obviously, for any value of p

$$\max \left\{ \begin{array}{l} \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell \\ \max_b \underset{\ell}{E} \pi(p,\ell,b) \end{array} \right\} \geq \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell$$

$$\max \left\{ \begin{array}{l} \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell \\ \max_b \underset{\ell}{E} \pi(p,\ell,b) \end{array} \right\} \geq \max E \pi(p,\ell,b)$$

Therefore,

$$\underset{p}{E} \max \left\{ \begin{array}{l} \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell \\ \max_b \underset{\ell}{E} \pi(p,\ell,b) \end{array} \right\} \geq \underset{p}{E}[\underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell$$

$$\underset{p}{E} \max \left\{ \begin{array}{l} \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_\ell \\ \max_b \underset{\ell}{E} \pi(p,\ell,b) \end{array} \right\} \geq \underset{p}{E}[\max_b \underset{\ell}{E} \pi(p,\ell,b)]$$

Subtracting $\max_b \underset{p}{E} \underset{\ell}{E} \pi(p,\ell,b)$ from both sides of these inequalities yields

$$V_p(K_\ell) \geq V_p^R(K\ )$$

$$V_p(K_\ell) \geq V_p^N$$

Thus for any value of K_ℓ , we are willing to pay at least as much to learn p , with an option to learn ℓ afterwards, as we would to learn p only, or to learn p and ℓ simultaneously. This result is illustrated in Fig. 2.8. The curve A-B-C-D for $V_p(K_\ell)$ is an upper bound for the straight line through C and D corresponding to V_p^N , and for the line A-E corresponding to $V_p^R(K_\ell)$.

In exactly the same way it can be shown that

$$V_\ell(K_p) \geq V_\ell^R(K_p)$$

$$V_\ell(K_p) \geq V_\ell^N$$

In Fig. 2.8 this means that the curve G-F-H-E corresponding to $V_\ell(K_p)$ must lie to the right of the straight line passing through G and F corresponding to V_ℓ^N , and the line A-E corresponding to $V_\ell^R(K_p)$.

We have shown that the set of feasible prices for sequential information includes the sets of feasible prices for individual and simultaneous purchases. From Fig. 2.8 it is obvious that the set of feasible prices for sequential purchases of information can include pairs of prices that would not be considered feasible when only individual or simultaneous purchases are considered. The conditions necessary for such pairs of prices to exist will be discussed in Chapter 4, but intuitively we can see that they will exist whenever the knowledge of one random variable could affect our decision to pay to learn the other random variable. Whenever learning a piece of information could cause us to change our decision, the information becomes more valuable than it would be if we had no subsequent decision to make. Because we have a subsequent decision when we buy information sequentially, the

value of perfect information can be greater in that case than when the information is purchased individually or sequentially.

Decision Regions in the Price Diagram

Although Fig. 2.8 shows that we would be willing to buy perfect information for any pair of prices represented by a point below and to the left of the curve G-F-H-C-D, it does not tell us which state variable to buy first. We should pay for the piece of information that will cause the greatest increase in our expected profit in order to maximize our expected profit. Since the increase in expected profit is the difference between the value and the cost of the information, we should first buy sequential information about p when

$$V_p(K_\ell) - K_p > V_\ell(K_p) - K_\ell$$

$$\underset{p}{E} \max \left\{ \begin{array}{l} \underset{\ell}{E} \max\limits_{b} \pi(p,\ell,b) - K_\ell \\ \\ \max\limits_{b} \underset{\ell}{E}\, \pi(p,\ell,b) \end{array} \right\} - \max_{b} \underset{p}{E}\, \underset{\ell}{E}\, \pi(p,\ell,b) - K_p$$

$$> \underset{\ell}{E} \max \left\{ \begin{array}{l} \underset{p}{E} \max\limits_{b} \pi(p,\ell,b) - K_p \\ \\ \max\limits_{b} \underset{p}{E}\, \pi(p,\ell,b) \end{array} \right\} - \max_{b} \underset{p}{E}\, \underset{\ell}{E}\, \pi(p,\ell,b) - K_\ell$$

$$\left\{ \begin{array}{ll} 29/96 - K_\ell - K_p & : K_\ell \le 1/8 \\ 33/96 + (4/3)K_\ell\sqrt{2K_\ell} - 2K_\ell - K_p & : 1/8 < K_\ell < 1/2 \\ 1/96 - K_p & : 1/2 \le K_\ell \end{array} \right\}$$

$$> \left\{ \begin{array}{ll} 29/96 + (2/3)K_p\sqrt{2K_p} - K_p/2 - K_\ell & : K_p < 1/8 \\ \\ 27/96 - K_\ell & : K_p \ge 1/8 \end{array} \right\}$$

In addition we will only pay to learn p when $V_p(K_\ell) \geq K_p$. Similar inequalities can be written to show when we should first buy sequential information about ℓ, or first buy information about p and ℓ simultaneously. These inequalities yield the decision regions shown in Fig. 2.9.

Using Fig. 2.9 we can immediately determine our best response when we are offered perfect sequential information about p and ℓ at prices K_p and K_ℓ, respectively. Figure 2.9 contains regions in which we would only want to pay for one piece of information without deciding afterwards whether or not to buy additional information. However, this is just a special case of learning the information sequentially. As far as the initial decision is concerned, there are only three alternatives: learn the actual value of p, learn the actual value of ℓ, or do not buy any information. Another way to express the information in Fig. 2.9 is to define three mutually-exclusive and collectively-exhaustive sets of price pairs, Ω_0, Ω_p, and Ω_ℓ, such that when $(K_p, K_\ell) \in \Omega_p$ our best initial decision is to pay K_p to learn p. When $(K_p, K_\ell) \in \Omega_\ell$ our best initial decision is to pay K_ℓ to learn ℓ, and when $(K\ ,K\ ) \in \Omega_0$, our best initial decision is to refuse all information. Ω_0, Ω_p, and Ω_ℓ are called decision sets and they correspond to the decision regions in Fig. 2.9.

Figure 2.9 shows that, for this problem, the option of buying perfect information about p and ℓ simultaneously is completely dominated by the other alternatives. This occurs in spite of the fact that there are pairs of prices such that buying both pieces of information simultaneously is preferable to buying either piece of information individually.

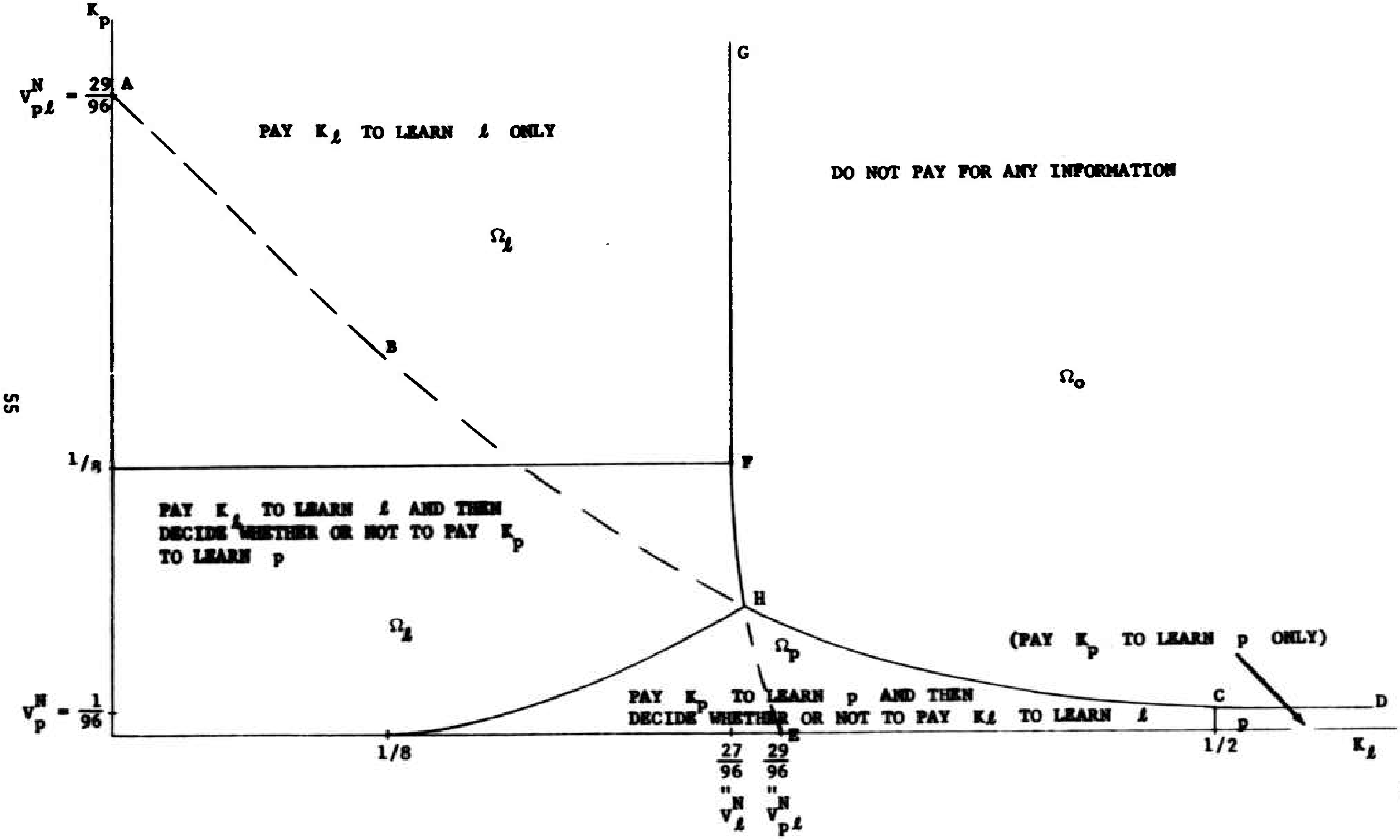


Figure 2.9. Decision regions for sequential information

If we compare Fig. 2.3 and Fig. 2.9, we can see that the same pair of prices can cause us to make a different initial decision when only individual or simultaneous information is available, than it can when sequential information is available. Figures 2.3 and 2.9 are shown superimposed in Fig. 2.10. If K_p and K_ℓ are represented by point A in Fig. 2-10, our best initial decision when offered individual or simultaneous information is to pay K_ℓ to learn ℓ . On the other hand, if we can buy information sequentially at the same prices, our best initial decision is to pay K_p to learn p . Similar comparisons can be made for other pairs of prices.

A Sequential Information Decision Tree

It has taken quite a bit of calculation to determine the decision regions shown in Fig. 2.9, even though our example is simple by comparison with many real-world decision problems. Fortunately all of this calculation is unnecessary if we simply want to know what to do when faced with a decision problem and a specific pair of prices for perfect information about the random variables. When the prices of perfect information are known, we can carry out a relatively simple calculation to find out which alternative maximizes the expected profit. For the bidding problem this calculation is equivalent to solving the decision tree shown in Figs. 2.11a and 2.11b.

Although the decision tree in Fig. 2.11 looks complicated, it is a straightforward calculation to solve the tree for the best decision and the resulting expected profit when K_p and K_ℓ are known. Thus for any pair of prices represented by a point in Fig. 2.9 we can solve for the

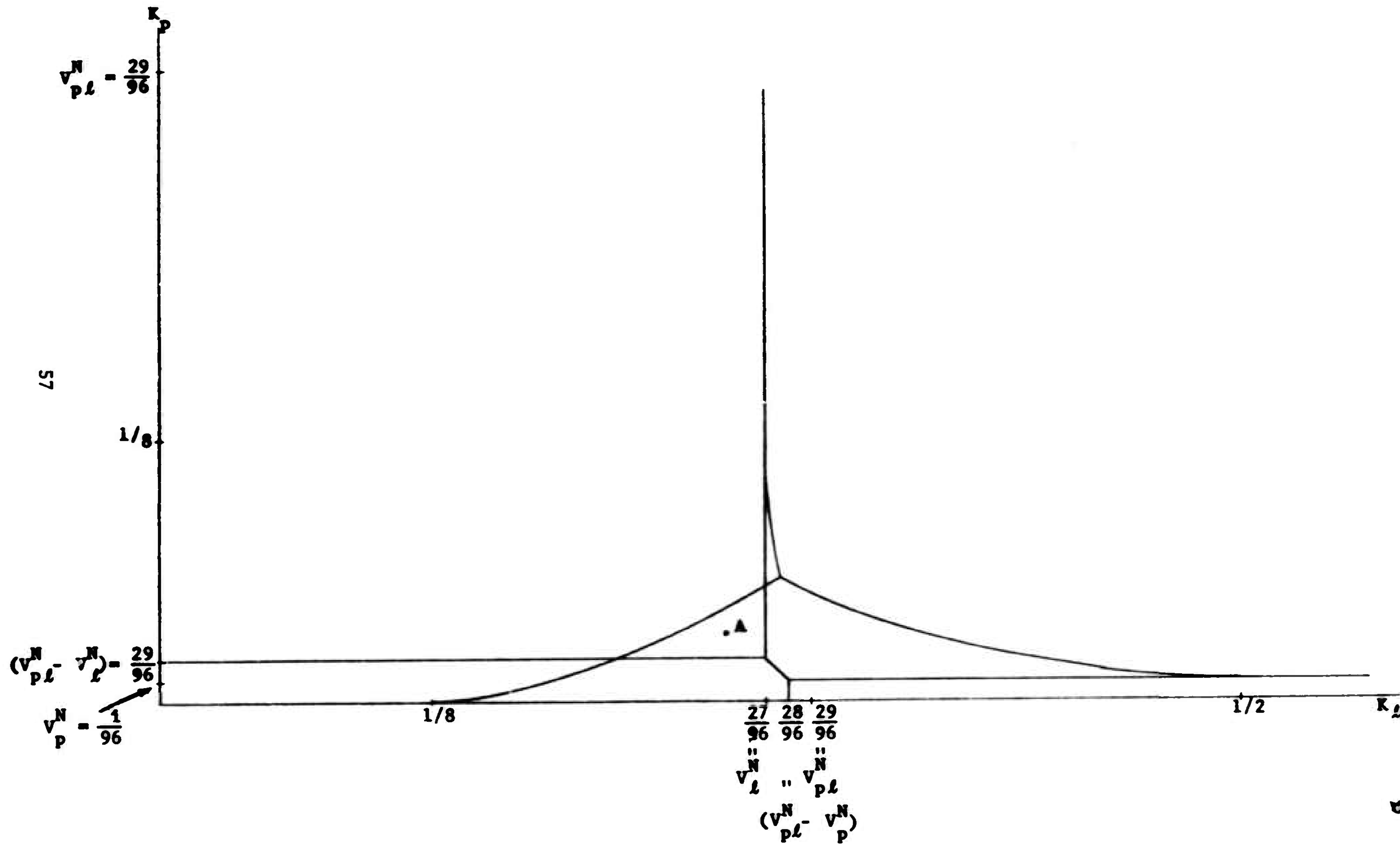


Figure 2.10. Comparison of decision regions with and without sequential information

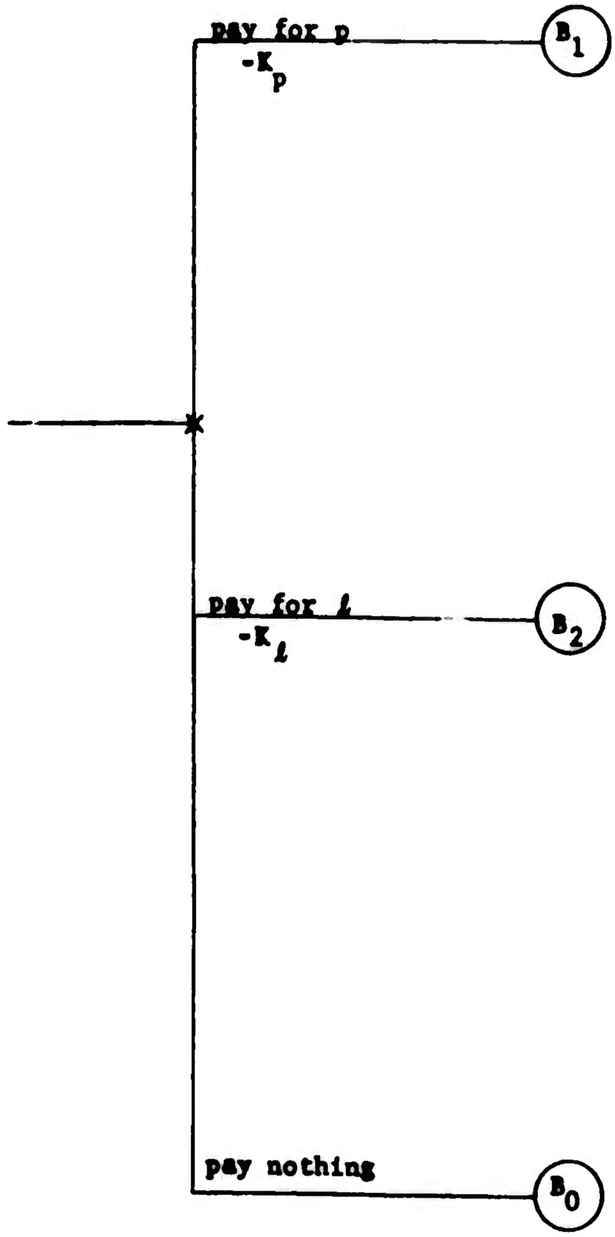


Figure 2.11a. Sequential information decision tree

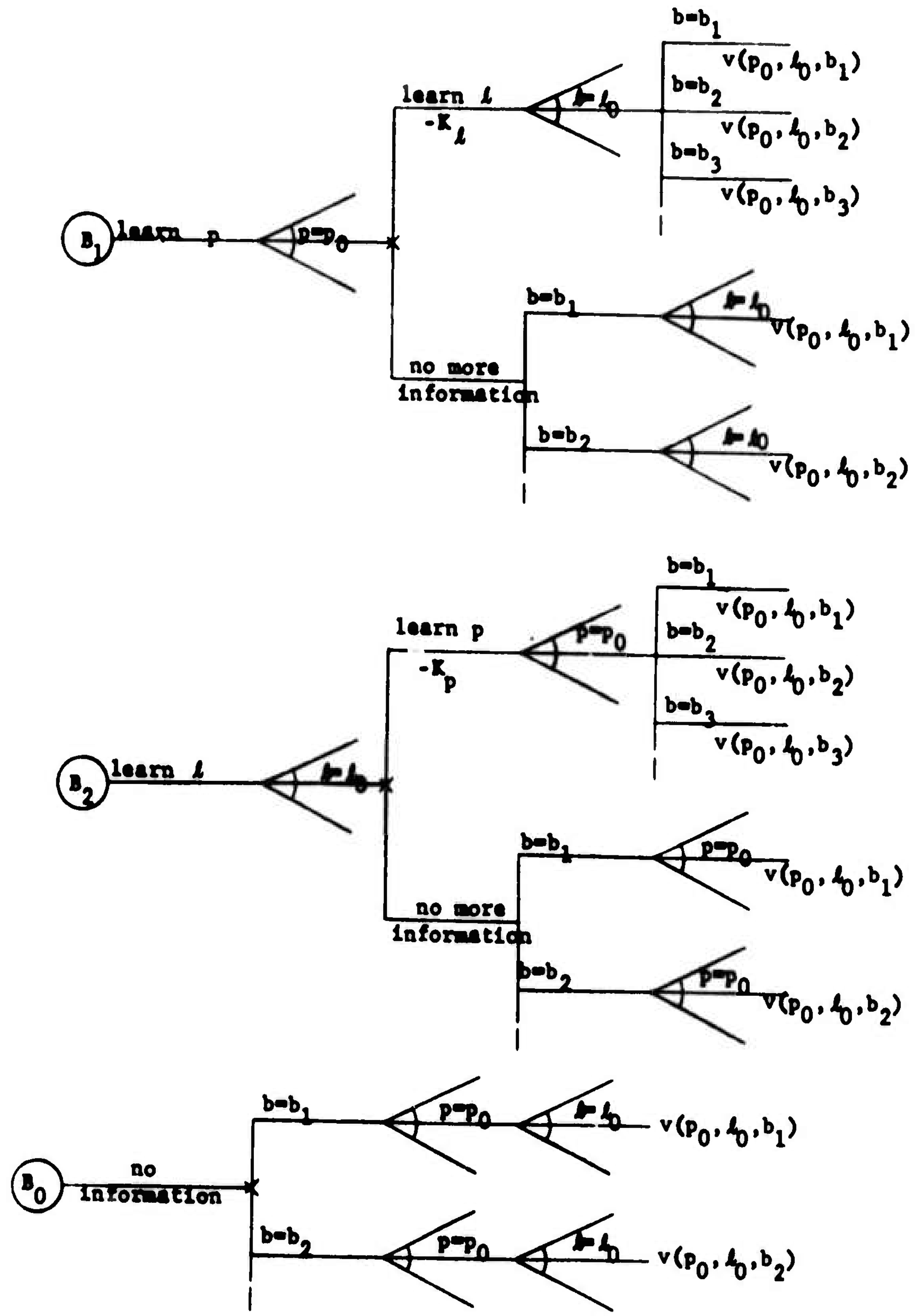


Figure 2.11b. Sequential information decision tree (continued)

optimal decision without determining all of the decision regions. By solving the decision tree for a number of different pairs of prices, we can determine the approximate boundaries of the decision regions in Fig. 2.9 without carrying out the previous calculations. For sequential information problems with many observables, this is often the only practical way to determine the decision regions. However, as we shall see, the decision tree can be used to characterize the decision regions when there are only two pieces of information available.

Characterizing the Decision Regions

If we consider those regions in Fig. 2.9 where our best initial decision is to buy information about p, we can see that we would be willing to pay a significantly greater amount for information about p when we subsequently have the option to buy perfect information about ℓ. Furthermore, when we restrict our attention to only those pairs of prices for which our best initial decision is to buy information about p, the amount we would be willing to pay to learn p reaches a maximum at the pair of prices corresponding to the point H. Although there are other pairs of prices for which we would be willing to pay more to learn p than we would at the pair of prices corresponding to the point H, we could increase our expected profit more by first paying to learn ℓ instead of p when we are offered one of these pairs of prices. Thus the maximum price that we would pay to learn p first is the K_p component of the point H. Exactly the same reasoning shows that the maximum amount that we would pay to learn ℓ first is the K_ℓ component of the point H.

The value of perfect information about a random variable is often

used to set an upper limit on the amount that we should pay to receive imperfect or partial information about the random variable. However, we have seen that the value of perfect information varies according to the price of perfect information about the other random variables. What, then, should we use for an upper bound on the value of imperfect information? The answer is the maximum value of perfect information, since it must be at least as large as the value of perfect information at any given pair of prices, which in turn thus must be at least as large as the value of imperfect information at the same pair of prices. It will be shown in the next chapter that the value of imperfect information also depends on the prices of the observables, and that the value of imperfect information taken sequentially can exceed the value of the same information when it is received alone.

In order to determine an upper bound for the value of imperfect information, we would like to have a procedure for finding the maximum value of perfect information (corresponding to the coordinates of the point H in Fig. 2.9) without going through all of the calculations required to determine the decision regions. Using the diagram in Fig. 2.9 we can demonstrate an iterative procedure for determining the coordinates of the point H to any degree of accuracy. This procedure will work for any sequential information problem with two observables.

Suppose we first determine the values of perfect information about p and ℓ, assuming in both cases that only one piece of information will be purchased. This gives us V_p^N and V_ℓ^N, the coordinates of point I in Fig. 2.12a. Now use the coordinates of point I as the prices, K_p and K_ℓ, in the decision tree branches of Fig. 2.11b to find the expected profits associated with branches B_1 and B_2. Subtracting the

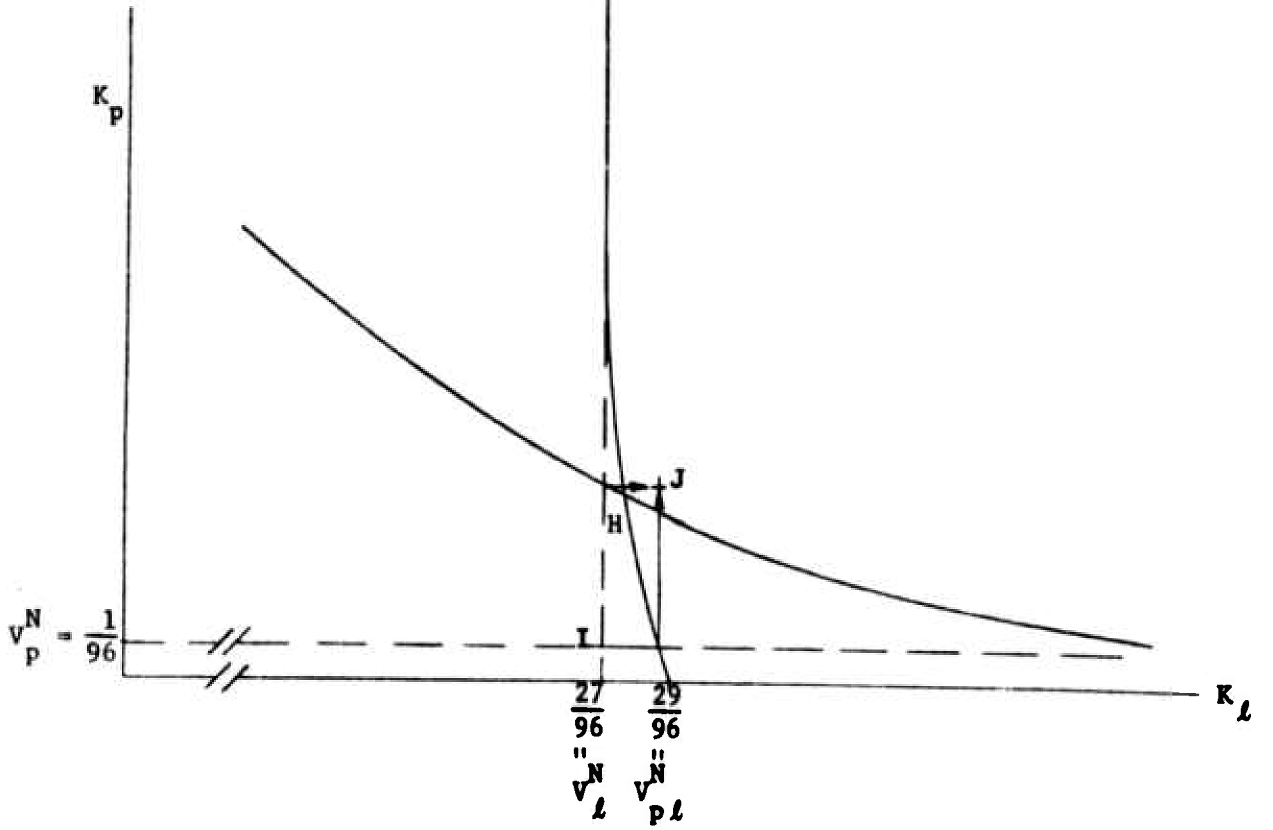


Figure 2.12a. Finding the maximum initial prices--first iteration

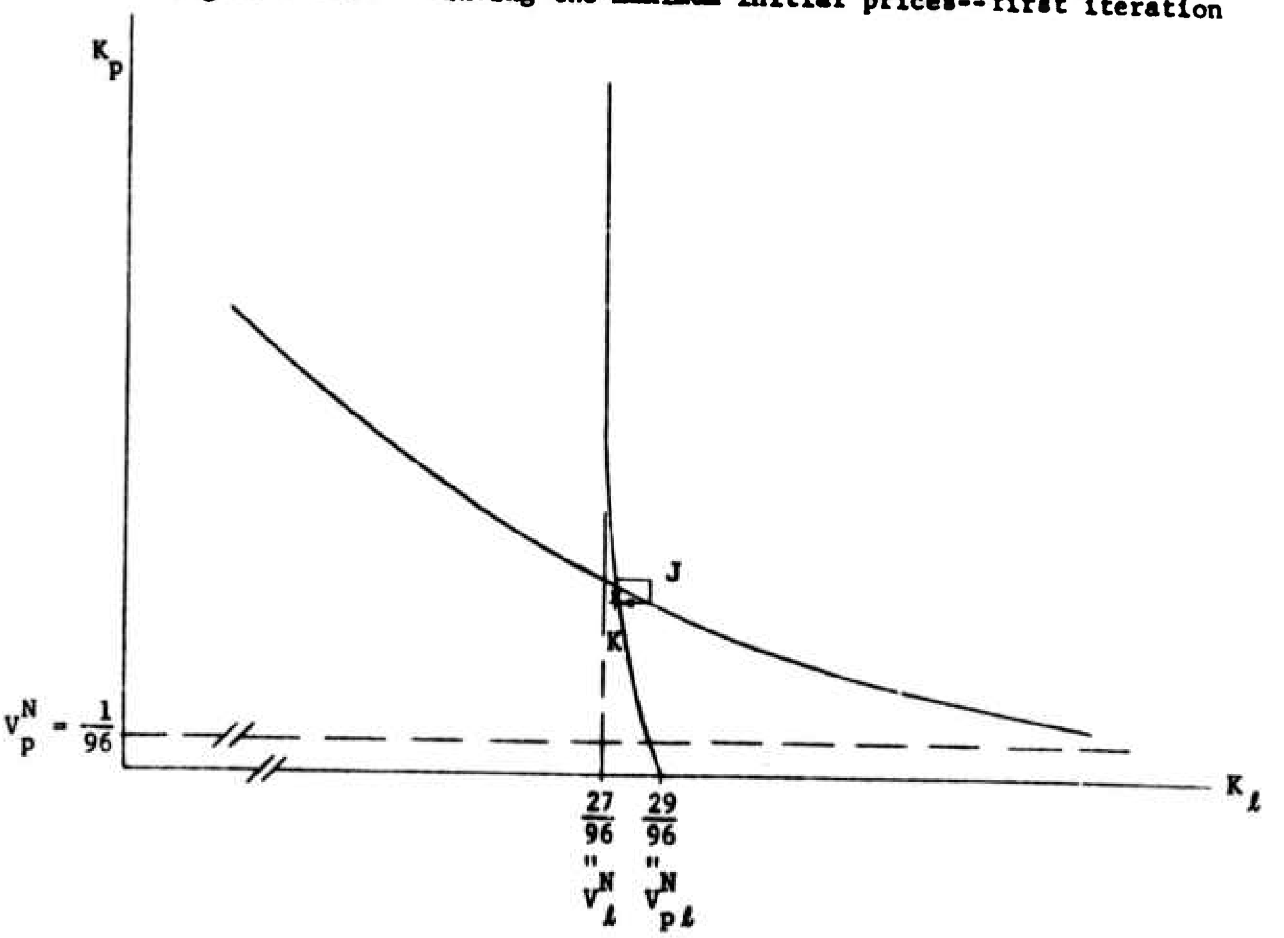


Figure 2.12b. Finding the maximum initial prices--second iteration

expected profit with no information (the expected profit associated with branch B_0) from the expected profits associated with branches B_1 and B_2) yields the values of sequential information about p and ℓ, when the cost of the information is given by V_p^N and V_ℓ^N. The cost of initially learning p or ℓ is shown in the tree in Fig. 2.11a, and is not included in the branches in Fig. 2.11b. Thus subtracting the expected profits associated with the branches in Fig. 2.11b will yield the desired values of information. It can be seen graphically in Fig. 2.12a that the resulting values of information are the coordinates of the point J.

Now repeat the procedure by using the coordinates of the point J as the prices, K_p and K_ℓ, in the decision tree in Fig. 2.11. Solving for the expected profits associated with branches B_1 and B_2 and then subtracting the expected profit associated with branch B_0 yields two new values of information. Figure 2.12b shows that these new values are the coordinates of the point K. The procedure can then be continued using the coordinates of the point K in the decision tree to get another point that is even closer to the point H.

Solving the decision tree in Fig. 2.11 repeatedly involves a considerable amount of calculation, so it is doubtful that anyone would want to carry this procedure through many steps. Fortunately it is not necessary to do so. The first iteration using V_p^N and V_ℓ^N in the decision tree is all that is needed to determine an upper bound for the value of perfect or imperfect information. It will be shown in Chapter 4 that for any sequential information problem with two observables the point J must always lie above and to the right of the point H if it does not

coincide with the point H. Thus the coordinates of the point J must be at least as large as the coordinates of the point H, which in turn must be at least as large as the values of perfect and imperfect information. Therefore, the coordinates of the point J are an upper bound for the value of sequential information about the random variables.

If there are more than two observables, the iterative procedure need not converge to a point whose coordinates are the maximum amounts that we would pay initially for sequential information about the observables. In fact, such a point need not exist. Furthermore the values of information determined from the appropriate decision tree, with the cost of the i^{th} observable set equal to $V^N_{y_i}$, need not be upper bounds for the maximum value of sequential information.

However, even though we cannot use the procedure discussed above to find the desired upper bounds when there are more than two observables, we can find a very simple upper bound for the maximum values of any number of observables by solving the decision tree with all of the prices set equal to zero. This corresponds to starting with a pair of prices represented by a point at the origin in Fig. 2.12. It is easy to see that solving any sequential-information decision tree with all of the prices set equal to zero will result in the value of any observable being equal to the value of learning all of the observables simultaneously. No matter which observable we decide to learn first, we will subsequently decide to learn all of the other observables because they are free. Thus the value of learning all of the observables simultaneously is an upper bound for the value of learning any of the observables sequentially.

For the biding problem this means that

$$V_{p\ell}^{N} \geq V_p(K_\ell) \quad \text{and} \quad V_{p\ell}^{N} \geq V_\ell(K_p)$$

for all K_p and K_ℓ. This is easy to verify graphically using Fig. 2.8.

Summary

For the bidding problem discussed above, we have found that the value of perfect information about one state variable is a function of the price of perfect information about the other state variable. We also found that as the price of information about one state variable declines, the value of information about the other increases to a level greater than that achieved by learning the second state variable alone.

The Euclidean space with the prices of the observables along the axes (the "price diagram") can be divided into decision regions which indicate the best initial decision. For any given pair of prices, the best initial decision, the expected profit, and the value of sequential information can be found by solving the appropriate decision tree. The decision tree can also be used to find the approximate boundaries of the decision regions, or, when there are only two observables, to find the maximum prices that we would pay initially to learn the observables.

CHAPTER 3

SEQUENTIAL IMPERFECT INFORMATION WITH ADDITIVE, CERTAIN PRICES: A WEATHER FORECASTING PROBLEM

The preceding chapter dealt with the possibility of buying perfect information. In this chapter we investigate the case of imperfect information using a simple example to illustrate the relevant concepts. The example addresses the problem of when to pay for imperfect weather forecasts while trying to decide if an activity should be held indoors or outdoors. In this problem we have the opportunity to buy two conditionally independent pieces of information, the weather forecasts, about a single state variable that describes the weather. The procedures developed in this chapter can be used when the observables are not conditionally independent, but the example is easier to follow and the results are clearer if we assume conditional independence. For simplicity, the example only deals with the possibility of buying two observables but the analysis applies equally well when there are more than two pieces of information available. Utility and risk aversion are not considered, but it is simple to extend the following calculations to include these concepts. The prices of the observables are assumed to be additive and certain.

The Weather Forecasting Problem

Consider a situation where we are trying to decide whether to have a certain activity take place indoors or outdoors. If it rains we would like to have the activity indoors; if there is fair weather we would like

to have it outdoors. We assume that the weather will either be fair or rain. Since we do not want to stage the activity in the wrong location we will have to move it if we do not forecast the weather correctly. Once the plans have been made and the activity has been set up, it will cost one unit (which could be a dollar, a million dollars, or any other amount) to move it to another location. We are being paid one unit to set up the activity so our net profit, as a function of the location we choose and the weather, is given by the following table:

Location	Weather	Profit
Indoors	Rain	One
Indoors	Fair	Zero
Outdoors	Rain	Zero
Outdoors	Fair	One

Our objective is to maximize our expected profit.

We can describe the state of the weather in terms of a discrete random variable. Let x be a random variable that equals one if it rains and zero if it does not rain. Similarly, we can describe our decision in terms of a discrete control variable c. Setting c equal to zero is equivalent to setting up the activity outdoors, and setting c equal to one is equivalent to setting up the activity indoors. Since we will have a net gain of one unit if we do not have to move the activity at the last minute, our profit function is

$$\pi(x,c) = \left\{ \begin{array}{ll} 1 : & c = x \\ 0 : & \text{otherwise} \end{array} \right\}$$

Since the random variable x appears in the profit function, x is a

state variable. In this problem we will not have an opportunity to observe the actual value of x before we make our decision. Therefore x is not an observable.

We do not know for certain if it will rain, but we have the option of paying one or more weather forecasting companies to learn what they think the weather will be. Assume that after thinking about the problem, but before buying a forecast, we decide that rain is twice as likely as fair weather. Thus the probability mass function for the random variable x is

$$\{x|\mathcal{E}\} = \begin{Bmatrix} 1/3 : x = 0 \\ 2/3 : x = 1 \\ 0 \;\; : \text{otherwise} \end{Bmatrix}$$

This mass function is shown in Fig. 3.1.

After investigating the reputation of the two weather forecasting companies, we have decided that they are equally reliable and that they are correct 60% of the time. In other words on days when it actually rained both companies predicted rain 60% of the time. The remaining 40% of the time they incorrectly forecast fair weather. Similarly on days when the weather was fair, both companies forecast fair weather 60% of the time and rain 40% of the time. Although the two companies have the same percentage of correct forecasts, their forecasts for any given day do not necessarily agree. In fact, our investigation indicates that, if we knew what the weather was on a certain day, knowing one forecast would not help us guess the other forecast.

We can describe the information contained in the weather forecasts in terms of two discrete random variables, y_1 and y_2. Let y_1 equal

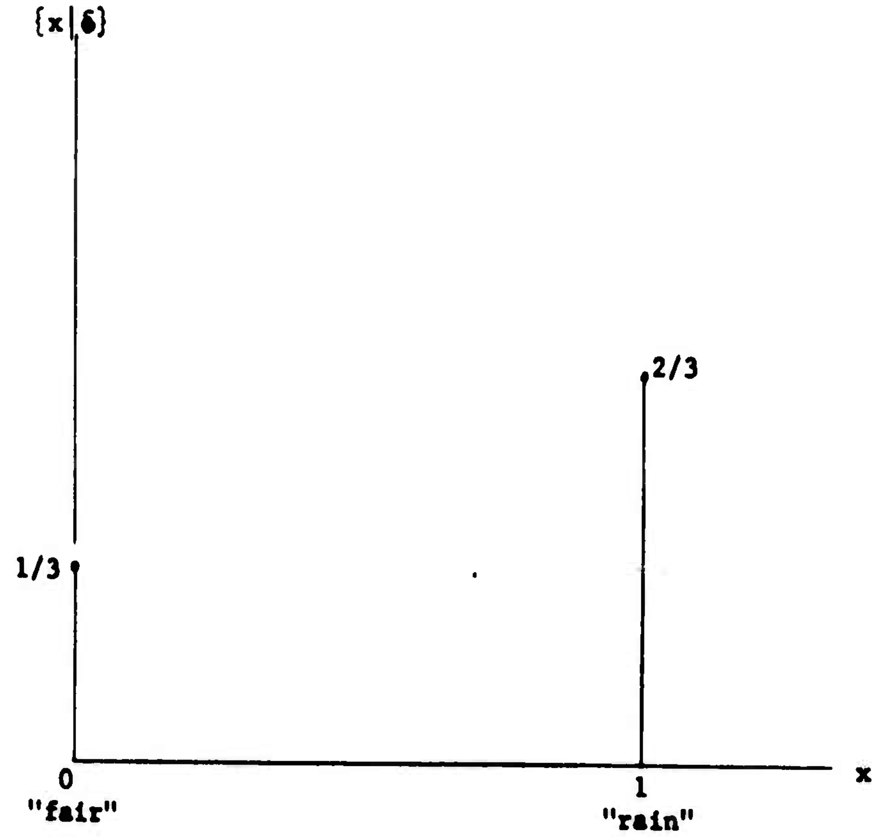


Figure 3.1. Probability mass function for random variable describing the weather

one if the first company predicts that it will rain when the activity is set up, and zero if that company predicts fair weather. Similarly, y_2 is defined to be one if the second company forecasts rain and zero otherwise. Based on our investigation, we can describe our state of information about y_1 and y_2 in terms of the following conditional probability mass functions:

$$\{y_i|x,\mathcal{E}\} = \begin{cases} 6/10 : x = 0\ , & y_i = 0 \\ 4/10 : x = 0\ , & y_i = 1 \\ 4/10 : x = 1\ , & y_i = 0 \\ 6/10 : x = 1\ , & y_i = 1 \\ 0 \quad : \text{otherwise} & \end{cases} \qquad (i = 1,2)$$

This mass function is shown in Fig. 3.2. y_1 and y_2 are called observables since we have an opportunity to learn their actual values.

<u>The Value of Individual, Simultaneous, and Sequential Information</u>

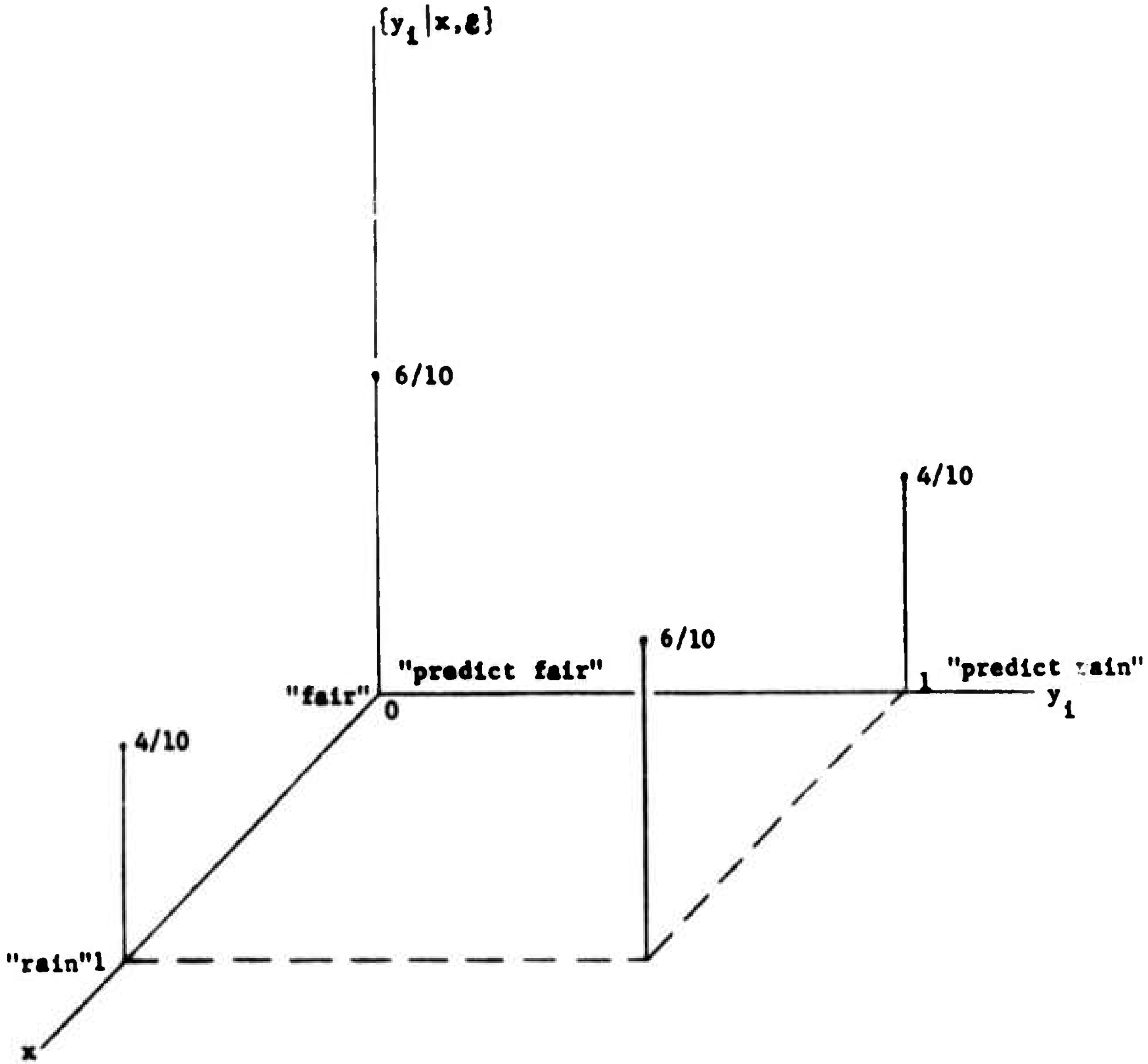
With the problem formulated in this manner, we can determine the value of the two weather forecasts when we learn them individually, simultaneously, and sequentially. Since the procedure for doing so is essentially similar to that shown in the previous chapter, we will only state the results here. The details of the derivation are given in Appendix B.

The value of learning the first forecast by itself is

$$V_{y_1}^N = (\underset{y_1}{E} \max_c \underset{x}{E} - \max_c \underset{x}{E})\ \pi(x,c) = 0$$

Similarly the value of learning the second forecast by itself is

$$V_{y_2}^N = (\underset{y_2}{E} \max_c \underset{x}{E} - \max_c \underset{x}{E})\ \pi(x,c) = 0$$



(i = 1,2)

Figure 3.2. Conditional probability mass function showing relation between weather and forecast

The value of learning both forecasts simultaneously is

$$V^N_{y_1 y_2} = (\underset{y_1}{E}\ \underset{y_2}{E} \max_c \underset{x}{E} - \max_c \underset{x}{E})\ \pi(x,c) = 1/75$$

The value of learning the first forecast when we have an option to pay K_{y_2} for the second forecast is

$$V_{y_1}(K_{y_2}) = \underset{y_1}{E} \max \left\{ \begin{array}{l} \underset{y_2}{E} \max_c \underset{x}{E}\ \pi(x,c) - K_{y_2} \\ \\ \max_c \underset{x}{E}\ \pi(x,c) \end{array} \right\} - \max_c \underset{x}{E}\ \pi(x,c)$$

$$= \left\{ \begin{array}{lll} 1/75 - (7/15)\ K_{y_2} & : & K_{y_2} < 1/35 \\ \\ 0 & : & K_{y_2} \geq 1/35 \end{array} \right\}$$

The value of learning the second forecast when we have an option to pay K_{y_1} for the first forecast is

$$V_{y_2}(K_{y_2}) = \underset{y_2}{E} \max \left\{ \begin{array}{l} \underset{y_1}{E} \max_c \underset{x}{E}\ \pi(x,c) - K_{y_1} \\ \\ \max_c \underset{x}{E}\ \pi(x,c) \end{array} \right\} - \max_c \underset{x}{E}\ \pi(x,c)$$

$$= \left\{ \begin{array}{lll} 1/75 - (7/15)\ K_{y_1} & : & K_{y_1} < 1/35 \\ \\ 0 & : & k_{y_1} \geq 1/35 \end{array} \right\}$$

The price diagram for this problem is shown in Fig. 3.3. Since the value of learning either forecast by itself is zero, there are no pairs of prices in Fig. 3.3 such that we would buy either observable by itself. We are willing to buy both forecasts simultaneously for any pair of prices represented by a point below and to the left of the line A-B. (Actually we would always have a higher expected profit if we

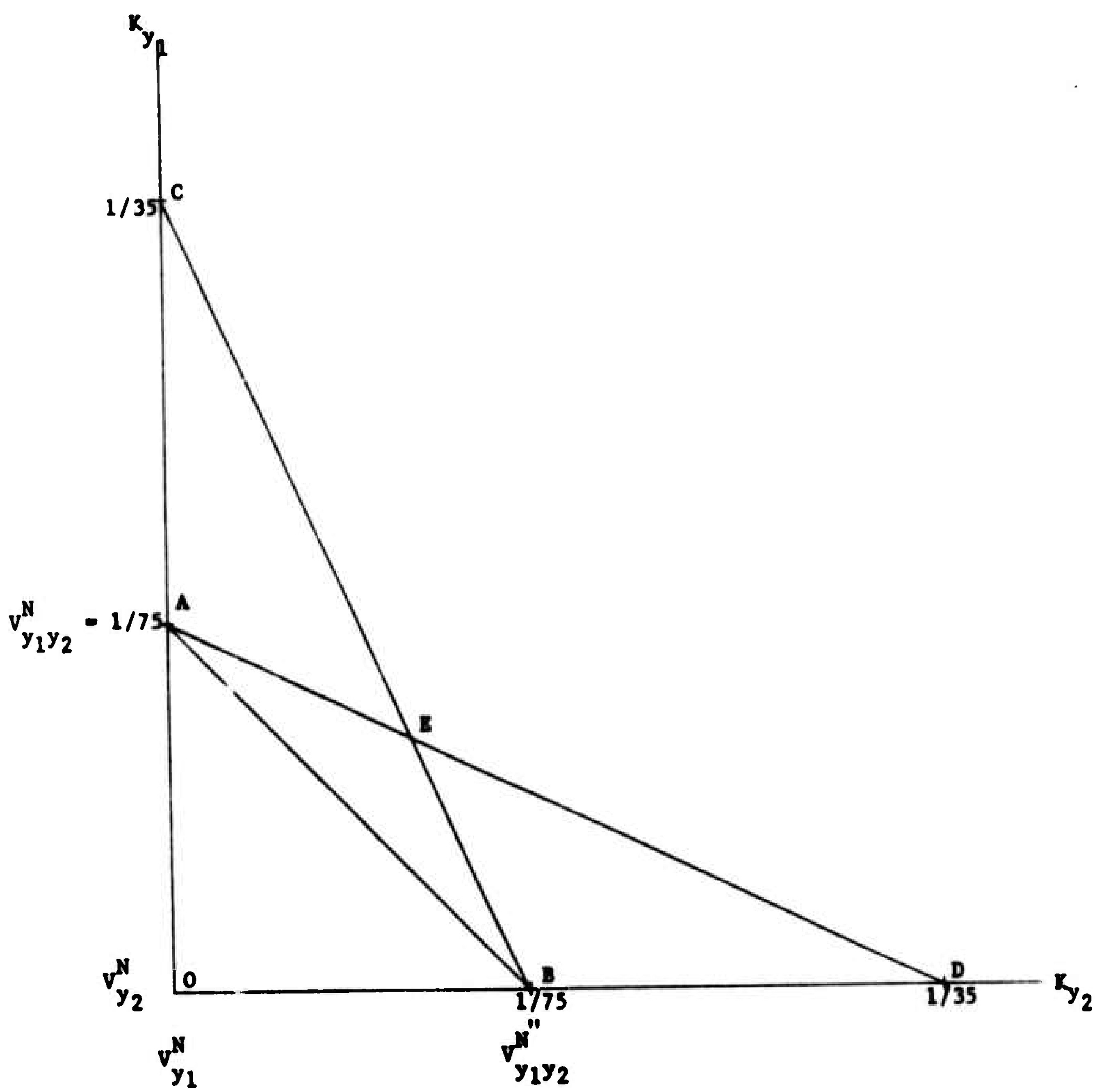


Figure 3.3. Price diagram with sequential information

purchased the forecasts sequentially instead of simultaneously, but the idea here is that buying both observables simultaneously is preferable to not buying any information.) We are willing to pay K_{y_1} for the first forecast, with an option to pay K_{y_2} for the second forecast, for any pair of prices below the line A-D. We are willing to pay K_{y_2} for y_2, with an option to pay K_{y_1} for y_1, for any pair of prices to the left of the line C-B. Therefore we are willing to buy at least one of the forecasts for any pair of prices below or to the left of the boundary C-E-D.

As in the case of perfect information, we can divide the price diagram into decision regions by comparing the increase in expected profit associated with each of the ways we can buy the information. If we do so, we find the decision regions shown in Fig. 3.4. The decision regions are considerably different than the ones we would find if the imperfect information were not available sequentially. In that case, the only decision region such that we would buy information is triangle O-A-B in Fig. 3.3.

The point of this simple example is that we deal with perfect and imperfect sequential information in the same way, with only minor changes in the algebra used to describe the value of information. We can determine the value of sequential imperfect information about any observable, and we can show that it must exceed or equal the value of the same information learned by itself (see Appendix B).

A Sequential Information Decision Tree

Using Fig. 3.4 we can determine our best response when we are

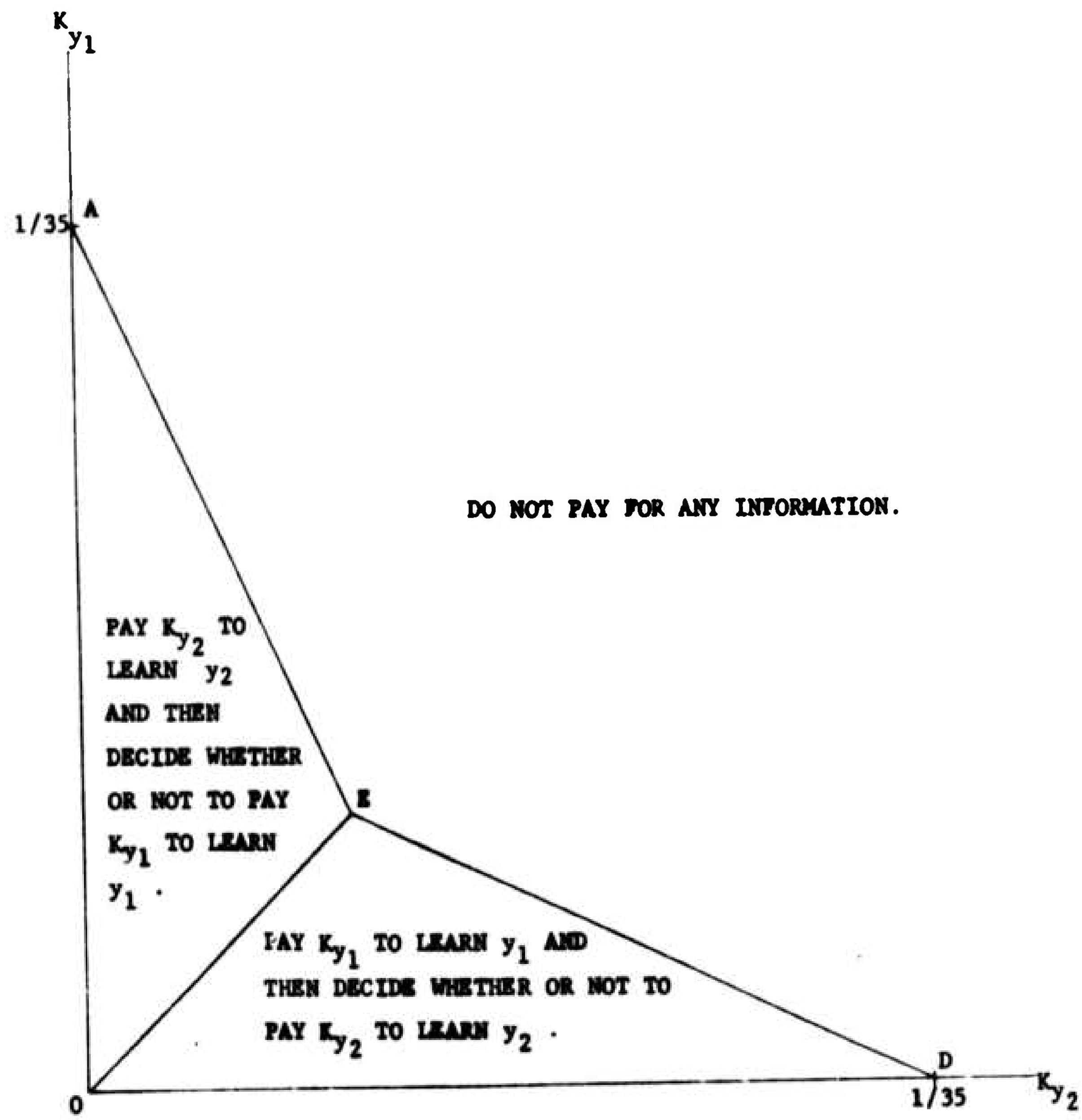


Figure 3.4. Decision regions with sequential information

offered the two weather forecasts at prices K_{y_1} and K_{y_2}. However, if we have a specific pair of prices and want to know what our best decision is, we do not need to carry out all of the calculations necessary to determine the decision regions. When the prices are known, we can carry out a relatively simple calculation to find the optimal decision and the corresponding expected profit. For the weather forecasting problem this is equivalent to solving the decision tree shown in Fig. 3.5.

By solving the decision tree if Fig. 3.5 we can find the best decision associated with a number of different points in the price diagram in Fig. 3.4. In this way we can determine the approximate boundaries of the decision regions without carrying out all of the calculations necessary to describe the decision regions algebraically. For sequential information problems with many observables, this is often the only practical way to determine the decision regions.

When there are only two observables we can use the decision tree in Fig. 3.5 to characterize the decision regions. We can characterize the decision regions by bounding the maximum price that we would be willing to pay for each observable when our best initial decision is to buy that observable. The procedure for doing this is similar to the one for perfect information, discussed in Chapter 2, and it is carried out for the weather forecasting problem in Appendix B.

The Relationship Between Perfect and Imperfect Sequential Information Problems

We have found that we can deal with sequential imperfect information in much the same way that we deal with sequential perfect information. This is not really surprising since imperfect information can be

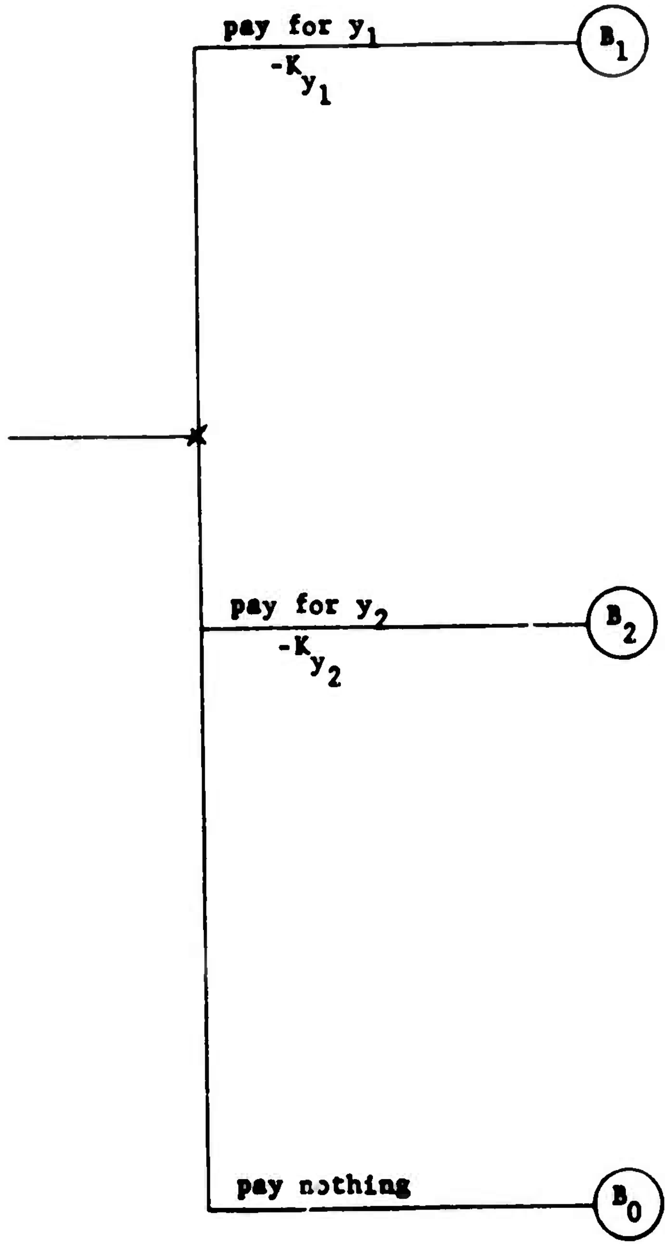


Figure 3.5a. Sequential information decision tree

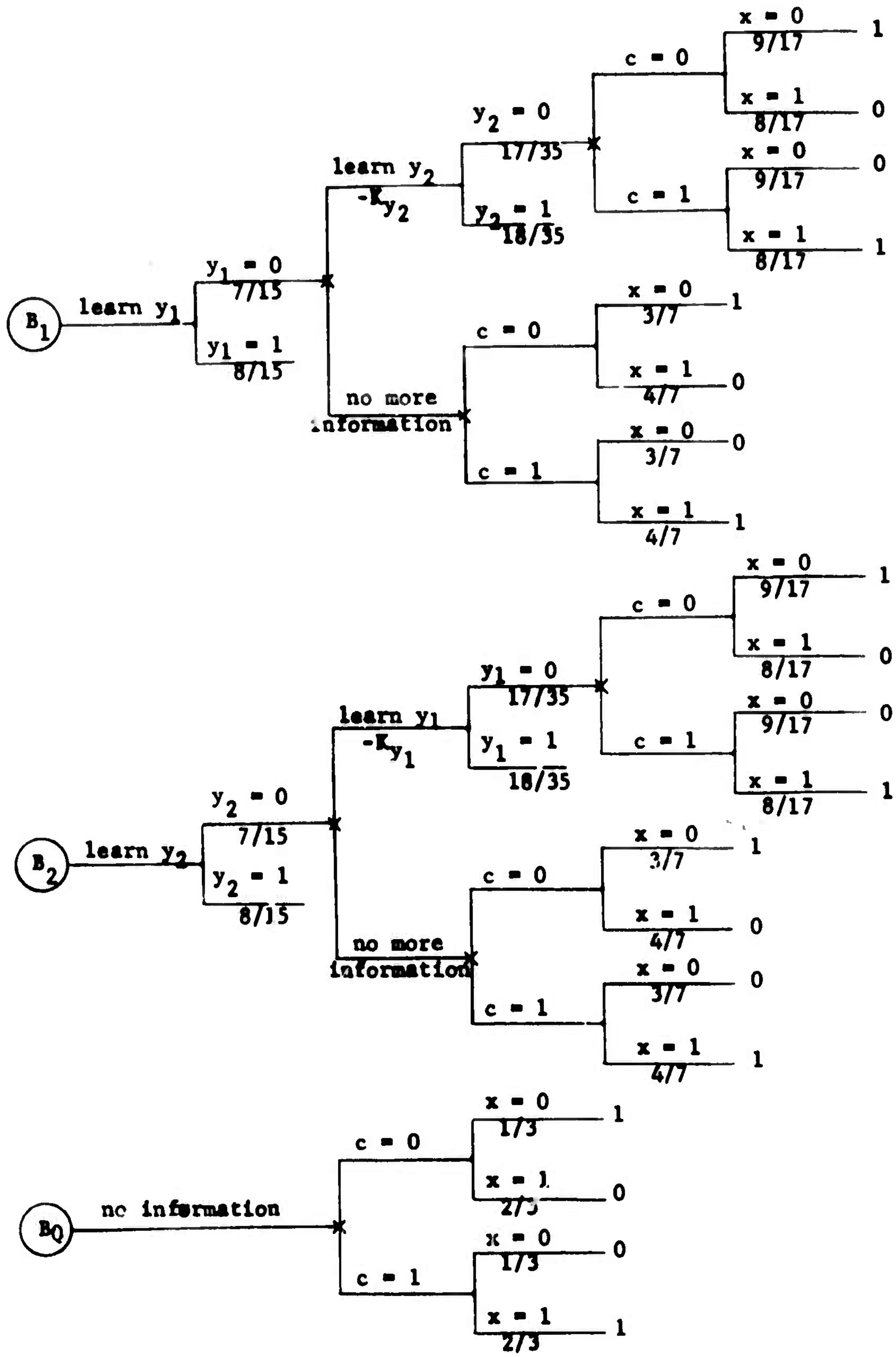


Figure 3.5b. Sequential information decision tree (continued)

viewed as perfect information about something other than the state variables. In fact, it is possible to formulate all sequential information problems with additive, certain prices as follows. Let $(x_1,\ldots,x_m)$ be the state variables upon which the profit depends. The profit also depends on a control variable c which may be a vector. Thus,

$\pi(x_1,\ldots,x_m,c)$ = profit function that depends on state and control variables

Assume we have an opportunity to learn any one of a set of random variables $(y_1,\ldots,y_n)$ called observables. The cost of learning y_i is K_{y_i} , where $i = 1,\ldots,n$. Some of the observables may be the same as some of the state variables, in which case we have an opportunity to buy perfect information, but this need not be the case. Even if we do not have perfect information, we must know how the observables are related to the state variables. These relationships can be deterministic or probabilistic.

Given this formulation we can find the value of learning an observable individually or sequentially, and we can find the value of learning several observables simultaneously, regardless of whether the observables represent perfect or imperfect information. For example, if we are to be given only the value of y_i and we do not expect to receive any additional information, our expected profit is

$$\underset{y_i}{E} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m)$$

If y_i represents imperfect information, as in the weather forecasting problem, this expression does not need to be simplified. However, if y_i is equal to x_j , so that y_i represents perfect information, then we

can simplify as follows:

$$\underset{y_i}{E} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m) = \underset{x_j}{E} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m)$$

$$= \underset{x_j}{E} \max_c \underset{x_1}{E} \cdots \underset{x_{j-1}}{E} \underset{x_{j+1}}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m)$$

The second $\underset{x_j}{E}$ operator can be dropped since it is just the identity operator when the expectation is conditioned on x_j .

The last expression above is the one we used in Chapter 2 for the expected profit when we were given perfect information about x_j . It is easy to show that all of the other equations used in Chapter 2 for perfect information can be expressed in terms of observables $(y_1,\ldots,y_n)$ and state variables $(x_1,\ldots,x_m)$. Thus this formulation can be used to deal with both perfect and imperfect information. However, it is unnecessarily complicated then we are only interested in the value of perfect information.

Summary

We have found that for the weather forecasting problem, the value of one piece of imperfect information is a function of the price of the other piece of information. We also found that the value of information received sequentially can exceed the value of the same information received individually. As in the case of perfect sequential information, we can divide the price diagram into decision regions that show the best initial information-purchasing decision. For any given pair of prices for the two weather forecasts we can determine the optimum initial decision by solving the appropriate decision tree. The decision tree can be

used to find the approximate boundaries of the decision regions. The results derived in this chapter for sequential imperfect information are similar to those found previously for sequential perfect information. In fact, we can use the same formulation for both types of problems. In this formulation, observables are distinguished from state variables even though they may represent the same thing.

CHAPTER 4

GENERAL PROPERTIES OF SEQUENTIAL INFORMATION PROBLEMS WITH ADDITIVE, CERTAIN PRICES

The basic concepts used in solving sequential information problems have been illustrated with examples in Chapters 2 and 3. However, the solution of these simple examples is not sufficient to demonstrate several general properties of sequential information problems with additive, certain prices. The purpose of this chapter is to generalize the results of Chapters 2 and 3, and also to prove some results that were only mentioned in passing during the discussion of the examples. For the reader who does not want to read all the proofs, the results are summarized at the end of the chapter.

For simplicity we will start by assuming that our objective is to maximize expected profit. This assumption allows us to ignore the effect of risk aversion, which only complicates the calculations without altering the general conclusions. After studying the case of an expected-profit decision maker, we will show that the conclusions also apply to a decision maker whose utility function satisfies the delta property.

The problem of sequential information can be formulated in general as it was at the end of Chapter 3. Our profit π is a function of a control variable c and a set of state variables $(x_1,\ldots,x_m)$. We have an opportunity to learn any of a set of observables $(y_1,\ldots,y_n)$ for prices $(K_{y_1},\ldots,K_{y_n})$, respectively.

One of the principal results to come out of the examples in Chapters

2 and 3 is that the value of sequential information about an observable is, in general, a function of the prices of all of the other observables. Thus,

$$V_{y_i} = V_{y_i}(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n})$$

For simplicity we sometimes write the expected value of sequential information about y_i as

$$V_{y_i}(K_{y_1},\ldots,K_{y_n})$$

where it is understood that V_{y_i} does not depend on K_{y_i}. It is easy to see that this concept is valid, in general, for any sequential information problem, although for a given problem V_{y_i} may not be a function of some (or even all) of the prices. The dependence of V_{y_i} on the prices of the observables can be seen in the complete decision tree for the problem. This tree is shown in Fig. 4.1. The value of learning y_i is equal to the difference between the expected profit associated with branch B_i (the branch where we first learn y_i) in Fig. 4.1b, and the expected profit associated with branch B_0 (the branch where we do not learn any information). The cost of initially learning y_i is shown in Fig. 4.1a, and is not included in the branches in Fig. 4.1b. Thus, subtracting the expected values associated with the branches in Fig. 4.1b yields the desired expected values of information.

The value of learning y_i sequentially when we are trying to maximize expected profit is given by the following algebraic expression

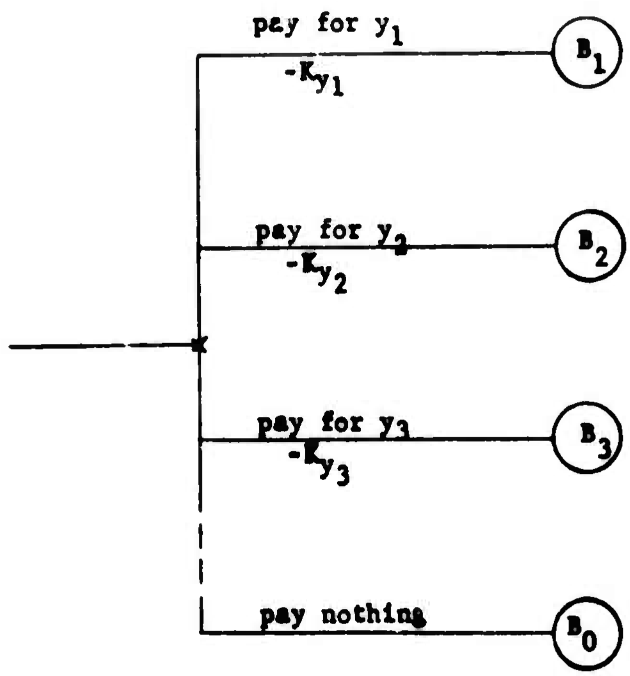


Figure 4.1a. Sequential information decision tree for an expected-profit decision maker

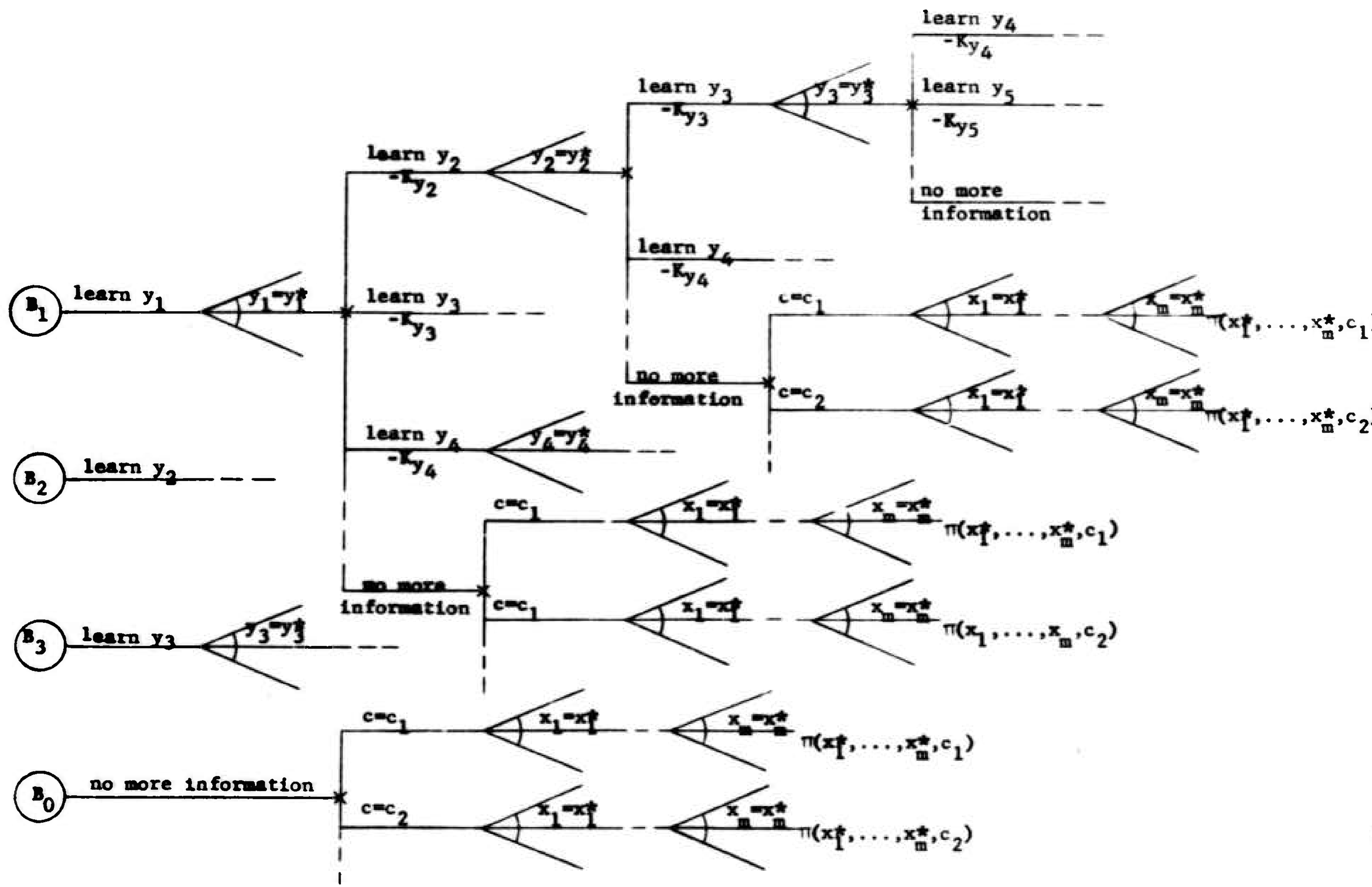


Figure 4.1b. Sequential information decision tree for an expected profit decision maker (continued)

$$V_{y_i}(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n}) =$$

$$\underset{y_i}{E}\ \max\left\{\begin{array}{l}\max\limits_{c}\ \underset{x_1}{E}\cdots\underset{x_m}{E}\ \pi(x_1,\ldots,x_m,c) \\ \max\limits_{j\neq i}\left(\underset{y_j}{E}\ \max\left\{\begin{array}{l}\max\limits_{c}\ \underset{x_1}{E}\cdots\underset{x_m}{E}\ \pi(x_1,\ldots,x_m,c) \\ \max\limits_{k\neq i,j}\ (\underset{y_k}{E}\ \max\{\ldots\} - K_{y_k})\end{array}\right\} - K_{y_j}\right)\end{array}\right\}$$

$$- \max_{c}\ \underset{x_1}{E}\cdots\underset{x_m}{E}\ \pi(x_1,\ldots,x_m,c)$$

In keeping with the notation introduced previously, each expectation is conditioned on the random variables that appear as subscripts of operators that are applied to the expectation in question. For example, all of the expectations in the expression above, except those in the final term that is subtracted from the rest of the expression, are conditioned on y_i. Since all of the costs except K_{y_i} appear on the right side of this equation, V_{y_i} must be a function of all of the costs except K_{y_i}. When we are trying to decide whether or not to pay for y_i, we must know all the costs, including K_{y_i}, because we will only buy the information when

$$K_{y_i} < V_{y_i}(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n})$$

The Relative Values of Individual, Simultaneous, and Sequential Information

It is easy to show that V_{y_i} must exceed or equal the value of individual information about y_i for any set of prices $(K_{y_1},\ldots,K_{y_n})$. The value of individual information is given by

$$V_{y_i}^N = \underset{y_i}{E} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi$$

Obviously, for any set of prices

$$\underset{y_i}{D} \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \max_{j \neq i} \left(\underset{y_j}{E} \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \max_{k \neq i,j} \left(\underset{y_k}{E} \max \{\ldots\} - K_{y_k} \right) \end{array} \right\} - K_{y_j} \right) \end{array} \right\}$$

$$\geq \underset{y_i}{D} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi$$

Taking the expected value with respect to y_i of both sides of this inequality and subtracting the expected profit when no information is purchased yields

$$V_{y_i}(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n}) \geq V_{y_i}^N$$

We can also show that the value of sequential information about y_i also exceeds the residual value of information about y_i when all of the information is purchased simultaneously. (We could also consider the residual value of learning y_i when we buy information simultaneously about all pairs, triples, etc., of observables. However, we will only consider the possibility of buying all of the observables simultaneously.) This requires a little more work, but the proof is essentially similar to the one above. The residual value of information about y_i is given by

$$V_{y_i}^R(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n}) = \prod_k (\underset{y_k}{E}) \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi$$

$$- \sum_{k \neq i} K_{y_k} - \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi$$

To save space we are using the following abbreviations:

$$\prod_k (D_{y_k}) = D_{y_1} \cdots D_{y_n} \quad \text{and} \quad \prod_{k \neq j} (D_{y_k}) = D_{y_1} \cdots D_{y_{j-1}} D_{y_{j+1}} \cdots D_{y_n}$$

We start the proof by noticing that for any $j_1 \neq i$

$$\prod_{k \neq j_1} (D_{y_k}) \max \left\{ \begin{array}{l} \max_c E_{x_1} \cdots E_{x_m} \pi \\ E_{y_{j_1}} \max_c E_{x_1} \cdots E_{x_m} \pi - K_{y_{j_1}} \end{array} \right\}$$

$$\geq \prod_{k \neq j_1} (D_{y_k}) E_{y_{j_1}} \max_c E_{x_1} \cdots E_{x_m} \pi - K_{y_{j_1}}$$

Let j_1^* be the value of j_1 that maximizes

$$\prod_{k \neq j_1} (D_{y_k}) E_{y_{j_1}} \max_c E_{x_1} \cdots E_{x_m} \pi - K_{y_{j_1}}$$

Since the inequality above holds for all values of j_1 not equal to i

$$\prod_{k \neq j_1^*} (D_{y_k}) \max \left\{ \begin{array}{l} \max_c E_{x_1} \cdots E_{x_m} \pi \\ E_{y_{j_1^*}} \max_c E_{x_1} \cdots E_{x_m} \pi - K_{y_{j_1^*}} \end{array} \right\}$$

$$\geq \prod_{k \neq j_1^*} (D_{y_k}) E_{y_{j_1^*}} \max_c E_{x_1} \cdots E_{x_m} > - K_{y_{j_1^*}}$$

Now take the expected value of both sides of this inequality with respect to y_{j_2}, given all of the other observables except $y_{j_1^*}$ --where j_2 does not equal j_1^* or i --and subtract $K_{y_{j_2}}$. This yields

$$\prod_{k \neq j_1^*, j_2} (D_{y_k}) \underset{y_{j_2}}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \underset{y_{j_1^*}}{E} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} \end{array} \right\} - K_{y_{j_2}}$$

$$\geq \prod_{k \neq j_1^*, j_2} (D_{y_k}) \underset{y_{j_2}}{E} \underset{y_{j_1^*}}{E} \max_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} - K_{y_{j_2}}$$

Therefore,

$$\prod_{k \neq j_1^*, j_2} (D_{y_k}) \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ E_{y_{j_2}} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \underset{y_{j_1^*}}{E} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} \end{array} \right\} - K_{y_{j_2}} \end{array} \right\}$$

$$\geq \prod_{k \neq j_1^*, j_2} (D_{y_k}) \underset{y_{j_2}}{E} \underset{y_{j_1^*}}{E} \max_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} - K_{y_{j_2}}$$

Now repeat the process. Let j_2^* be the value of j_2 that maximizes

$$\prod_{k \neq j_1^*, j_2} (D_{y_k}) \underset{y_{j_2}}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \underset{y_{j_1^*}}{E} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} \end{array} \right\} - K_{y_{j_2}}$$

Since the inequality above holds for all values of j_2 not equal to j_1^* or i :

$$\prod_{k\neq j_1^*, j_2^*} (D_{y_k}) \max \left\{ \begin{array}{l} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \\ \underset{y_{j_2^*}}{E} \max \left\{ \begin{array}{l} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \underset{y_{j_1^*}}{E} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} \end{array} \right\} - K_{y_{j_2^*}} \end{array} \right\}$$

$$\geq \prod_{k\neq j_1^*, j_2^*} (D_{y_k}) \underset{y_{j_2^*}}{E} \underset{y_{j_1^*}}{E} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} - K_{y_{j_2^*}}$$

Next we take the expected value of both sides of this inequality with respect to y_{j_3} given all of the other observables except $y_{j_1^*}$ and $y_{j_2^*}$ --where j_3 does not equal j_1^* , j_2^* , or i --and subtract $K_{y_{j_3}}$.

The process is repeated until all of the observables except y_i have been included in the inequality. At that point the inequality is

$$\underset{y_i}{D} \max \left\{ \begin{array}{l} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \\ \underset{y_{j_{n-1}^*}}{E} \max \left\{ \begin{array}{l} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \underset{y_{j_{n-2}^*}}{E} \max \{\ldots\} - K_{y_{j_{n-2}^*}} \end{array} \right\} - K_{y_{j_{n-1}^*}} \end{array} \right\}$$

$$\geq \underset{y_i}{D} \underset{y_{j_n^*}}{E} \underset{y_{j_{n-1}^*}}{E} \cdots \underset{y_{j_1^*}}{E} \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - K_{y_{j_1^*}} - \cdots - K_{y_{j_{n-1}^*}}$$

Since all of the indices, $(j_1^*, \ldots, j_{n-1}^*)$, must be different and none can be equal to i , the right side of this inequality can be written

$$\underset{y_i}{D} \prod_{k\neq i} (\underset{y_k}{E}) \max\limits_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - \sum_{k\neq i} K_{y_k}$$

To save space we are using the following abbreviation

$$\prod_{k\neq i} (\underset{y_k}{E}) = \underset{y_1}{E}, \ldots, \underset{y_{i-1}}{E}, \underset{y_{i+1}}{E}, \ldots, \underset{y_n}{E}$$

Using the definition of the starred indices, we have

$$\underset{y_i}{D} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \\ \max\limits_{j_{n-1} \neq i} \left(\underset{y_{j_{n-1}}}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \max\limits_{j_{n-2} \neq i, j_{n-1}} (\ldots) \end{array} \right\} - K_{y_{j_{n-1}}} \right) \end{array} \right\}$$

$$\geq \underset{y_i}{D} \prod_{k \neq i} (\underset{y_k}{E}) \max_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - \sum_{k \neq i} K_{y_k}$$

Taking the expected value with respect to y_i of boths sides of this inequality, and subtracting the expected profit when no information is purchased, yields

$$V_{y_i}(K_{y_1}, \ldots, K_{y_n}) \geq V^R_{y_i}(K_{y_1}, \ldots, K_{y_n})$$

for any set of prices $(K_{y_1}, \ldots, K_{y_n})$.

These inequalities agree with our intuition about the value of sequential information. When we buy a piece of information with an option to subsequently buy other observables, we have already been given the option of buying just one observable, or of buying all of the information. Thus the value of sequential information should be at least as large as the value of individual information and the residual value of information. We can always achieve the expected profit associated with individual and simultaneous information by making the proper set of sequential decisions. The same kind of reasoning shows that the value of sequential information of y_i must exceed the residual value of y_i , when we buy information simultaneously abour all pairs, triples, etc., of observables. Thus it is possible to prove a large number of inequalities of the form

$$V_{y_i} \geq V^N_{y_i y_j} - K_{y_j} \quad \text{and} \quad V_{y_i} \geq V^N_{y_i y_j y_k} - K_{y_j} - K_{y_k}$$

An interesting feature of the examples in Chapters 2 and 3 was that V_{y_i} could exceed both $V^N_{y_i}$ and $V^R_{y_i}$ for certain sets of prices. It is this feature that makes it desirable for us to buy sequential information at prices that would be unacceptable if the information were available only individually or simultaneously. Intuitively, we see that V_{y_i} can exceed both $V^N_{y_i}$ and $V^R_{y_i}$ whenever learning y_i can affect our decision to learn the other observables. If y_i can help us decide which pieces of information to buy, as well as how to set the control variable, then it should be worth more than it would if there were no subsequent information-purchasing decision.

It is easy to see that this intuitive answer is correct. Suppose that for some set of prices $(K_{y_1},\ldots,K_{y_n})$, learning y_i could affect our decision to learn the other observables. In other words, when we learn that y_i is equal to y_i^* we can maximize our expected profit by making a certain decision. This decision might be to learn a particular observable next, or it might be to not learn any more information. However, if we learn that y_i has a different value, y_i^{**}, we will be able to realize a greater expected profit by changing our decision.

Since the best information-purchasing decisions that result when we learn that y_i equals y_i^* or y_i^{**} are different, at least one of the decisions must be to buy additional information. Therefore we have

$$\underset{y_i}{D}\ \underset{y_j}{E}\ \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \\ \max_{k \neq i, j} \left(\underset{y_k}{E} \max \{\ldots\} - K_{y_k}\right) \end{array} \right\} - K_{y_j}$$

$$> \underset{y_i}{D} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi$$

for some j when y_i equals either y_i^* or y_i^{**}. Thus for the same conditions we have

$$\underset{y_i}{D} \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \\ \max_{j \neq i} \left(\underset{y_j}{E} \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \max_{k \neq i, j} \left(\underset{y_k}{E} \max \{\ldots\} - K_{y_k}\right) \end{array} \right\} - K_{y_j} \right) \end{array} \right\}$$

$$> \underset{y_i}{D} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E} \pi$$

Using our previous results, we can take the expected value of both sides of this inequality with respect to y_i and subtract the expected profit when no information is purchased to get

$$V_{y_i}(K_{y_1}, \ldots, K_{y_{i-1}}, K_{y_{i+1}}, \ldots, K_{y_n}) > V^N_{y_i}$$

On the other hand, the best information-purchasing decisions that result when we learn that y_i equals y_i^* or y_i^{**} cannot both lead to buying all of the remaining pieces of information regardless of what the information turns out to be. If that happened, the two decisions would lead to the same expected profit, and there would be no reason to change

the information-purchasing decision when y_i changes from y_i^* to y_i^{**}. Thus when the observables turn out to have certain values $(y_i, y_{j_1}, y_{j_2}, \ldots)$, and we make the corresponding optimum information-purchasing decisions, it must be true that

$$\underset{y_i}{D}\ \underset{y_{j_1}}{D}\ \underset{y_{j_2}}{D} \cdots \max_c \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi$$

$$> \underset{y_i}{D}\ \underset{y_{j_1}}{D}\ \underset{y_{j_2}}{D} \cdots \max_{k \neq i, j_1, j_2, \ldots} \left(\underset{y_k}{E} \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi \\ \max_{\ell \neq i, j_1, \ldots, k} \left(\underset{y_\ell}{E} \max \{\ldots\} - K_{y_\ell} \right) \end{array} \right\} - K_{y_k} \right)$$

$$\geq \underset{y_i}{D}\ \underset{y_{j_1}}{D}\ \underset{y_{j_2}}{D} \cdots \prod_{k \neq i, j_1, j_2, \ldots} \left(\underset{y_k}{E} \right) \max_c \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi - \sum_{k \neq i, j_1, j_2, \ldots} \left(K_{y_k} \right)$$

for some $j_1, j_2, \ldots$

when y_i equals y_i^* or y_i^{**}. Thus, for the same conditions we have

$$\underset{y_i}{D}\ \underset{y_{j_1}}{D}\ \underset{y_{j_2}}{D} \cdots \max \left\{ \begin{array}{l} \max_c \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi \\ \max_{k \neq i, j_1, j_2, \ldots} \left(\underset{y_k}{E} \max\{\ldots\} - K_{y_k} \right) \end{array} \right\}$$

$$> \underset{y_i}{D}\ \underset{y_{j_1}}{D}\ \underset{y_{j_2}}{D} \cdots \prod_{k \neq i, j_1, j_2, \ldots} \left(\underset{y_k}{E} \right) \max_c \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi - \sum_{k \neq i, j_1, j_2, \ldots} \left(K_{y_k} \right)$$

Now we can go through the same process which we used to show that $V_{y_i} \geq V^R_{y_i}$, except that this time all the inequalities are strict. We keep adding observables until we have

$$
\underset{y_i}{D} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \max\limits_{j_1 \neq i} \left(\underset{y_{j_1}}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi \\ \max\limits_{j_2 \neq i, j_1} \left(\underset{y_{j_2}}{E} \max\{\ldots\} - K_{y_{j_2}} \right) \end{array} \right\} - K_{y_{j_1}} \right) \end{array} \right\}
$$

$$
> \underset{y_i}{D} \prod_{k \neq i} \left(\underset{y_k}{E} \right) \max_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi - \sum_{k \neq i} K_{y_k}
$$

when y_i equals y_i^* or y_i^{**}

Using our previous results, we can take the expected value with respect to y_i of both sides of this inequality, and subtract the expected profit when no information is received, to get

$$
V_{y_i}(K_{y_1}, \ldots, K_{y_n}) > V^R_{y_i}(K_{y_1}, \ldots, K_{y_n})
$$

Thus for any given set of prices, V_{y_i} will exceed both $V^N_{y_i}$ and $V^R_{y_i}$ whenever learning y_i could affect our decision to learn the other observables.

Properties of the Value of Sequential Information

Now that we have shown that V_{y_i} is a function of the set of prices $(K_{y_1}, \ldots, K_{y_{i-1}}, K_{y_{i+1}}, \ldots, K_{y_n})$, we can determine some general properties of this function. More specifically, we can show that as any one of the prices increases, V_{y_i} cannot increase. In fact, it may decrease but, if the decision is based on maximizing expected profit, it cannot decrease by more than the amount that the price increases. This result follows from the form of the equation for V_{y_i} ,

$$V_{y_i} = \mathop{E}_{y_i} \max \left\{ \begin{array}{l} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} \pi \\[2ex] \max\limits_{j \neq i} \left(\mathop{E}\limits_{y_j} \max \left\{ \begin{array}{l} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} \pi \\[2ex] \max\limits_{k \neq i,j} \left(\mathop{E}\limits_{y_k} \max\{\ldots\} - K_{y_k} \right) \end{array} \right\} - K_{y_j} \right) \end{array} \right\}$$

$$- \max_{c} \mathop{E}_{x_1} \cdots \mathop{E}_{x_n} \pi$$

Each price, except K_{y_i} , appears in this expression, and each time the price is subtracted from some quantity. Thus we would expect the derivative of V_{y_i} with respect to one of the prices to be the expected value of terms that are either zero or minus one.

We can carry out the differentiation by using the fact that the derivative of the expected value is equal to the expected value of the derivative,

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \mathop{E}_{y_i} \frac{\partial}{\partial K_{y\alpha}} \max \left\{ \begin{array}{l} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} \pi \\[2ex] \max\limits_{j \neq i} \left(\mathop{E}\limits_{y_j} \max\{\ldots\} - K_{y_j} \right) \end{array} \right\}$$

$$= \mathop{E}_{y_i} \left\{ \begin{array}{l} 0 : \mathop{D}\limits_{y_i} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} \pi \geq \mathop{D}\limits_{y_i} \max\limits_{j \neq i} \left(\mathop{E}\limits_{y_j} \max\{\ldots\} - K_{y_j} \right) \\[2ex] \dfrac{\partial}{\partial K_{y_\alpha}} \max\limits_{j \neq i} \left(\mathop{E}\limits_{y_j} \max\{\ldots\} - K_{y_j} \right) : \text{otherwise} \end{array} \right\}$$

Let j^* be the value of j that maximizes

$$\mathop{D}_{y_i} \left(\mathop{E}_{y_j} \max \left\{ \begin{array}{l} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} \pi \\[2ex] \max\limits_{k \neq i,j} \left(\mathop{E}\limits_{y_k} \max\{\ldots\} - K_{y_k} \right) \end{array} \right\} - K_{y_j} \right)$$

and assume for a moment that j^* is unique. In this case the derivative becomes

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \underset{y_i}{E} \left\{ \begin{array}{l} 0 : \underset{y_i}{D} \ \underset{c}{\max} \ \underset{x_1}{E} \ \cdots \ \underset{x_n}{E} \geq \underset{y_i}{D} \ \underset{y_{j*}}{E} \ \max\{\ldots\} - K_{y_{j*}} \\ \\ \dfrac{\partial}{\partial K_{y_\alpha}} \ \underset{y_{j*}}{E} \ \max \left\{ \begin{array}{l} \underset{c}{\max} \ \underset{x_1}{E} \ \cdots \ \underset{x_n}{E} \ \pi \\ \underset{k \neq i,j}{\max} \ (\ldots) \end{array} \right\} - K_{y_{j*}} \quad : \text{otherwise} \end{array} \right\}$$

If $j*$ equals α, we have

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \underset{y_i}{E} \left\{ \begin{array}{l} 0 : \underset{y_i}{D} \ \underset{c}{\max} \ \underset{x_1}{E} \ \cdots \ \underset{x_n}{E} \ \pi \geq \underset{y_i}{D} \ \underset{y_\alpha}{E} \ \max\{\ldots\} - K_{y_\alpha} \\ -1 : \text{otherwise} \end{array} \right\}$$

The expected value of a quantity that is everywhere either 0 or -1 must lie in the closed interval bounded by these numbers. Therefore,

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \in [-1,0]$$

Now consider what happens to $\partial V_{y_i}/\partial K_{y_\alpha}$ as K_{y_α} increases. Let $\varphi_1(K_{y_\alpha})$ be the set of all y_i such that the condition in the equation for $\partial V_{y_i}/\partial K_{y_\alpha}$ is met:

$$\varphi_1(K_{y_\alpha}) = [y_i : K_{y_\alpha} \geq \underset{y_i}{D} \ \underset{y_\alpha}{E} \ \max\{\ldots\} - \underset{y_i}{D} \ \underset{c}{\max} \ \underset{x_1}{E} \ \cdots \ \underset{x_n}{E} \ \pi]$$

We can rewrite the derivative of V_{y_i} as follows:

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \underset{y_i}{E} \left\{ \begin{array}{l} 0 \ \ : y_i \in \varphi_1(K_{y_\alpha}) \\ \\ -1 : \text{otherwise} \end{array} \right\}$$

Assume that y_i^o belongs to $\varphi_1(K_{y_\alpha})$, and let K'_{y_α} be some price larger than K_{y_α},

$$K'_{y_\alpha} > K_{y_\alpha} \geq \underset{y_i^o}{D} \underset{y_\alpha}{E} \max\{\ldots\} - \underset{y_i}{D} \underset{c}{\max} \underset{x_1}{E} \cdots \underset{x_n}{E} \pi$$

Therefore, y_i^o belongs to $\varphi_1(K'_{y_\alpha})$ as well as $\varphi_1(K_{y_\alpha})$. Since this holds for any y_i^o that belongs to $\varphi_1(K_{y_\alpha})$

$$\varphi_1(K_{y_\alpha}) \subseteq \varphi_1(K'_{y_\alpha})$$

whenever

$$K'_{y_\alpha} > K_{y_\alpha}$$

So,

$$\frac{\partial V_{y_i}}{\partial K'_{y_\alpha}} = \underset{y_i}{E} \left\{ \begin{array}{l} 0 \ : y_i \in \varphi_1(K'_{y_\alpha}) \\ -1 : \text{otherwise} \end{array} \right\}$$

$$\geq \underset{y_i}{E} \left\{ \begin{array}{l} 0 \ : y_i \in \varphi_1(K_{y_\alpha}) \subseteq \varphi_1(K'_{y_\alpha}) \\ -1 : \text{otherwise} \end{array} \right\} = \frac{\partial V_{y_i}}{\partial K_{y_\alpha}}$$

In other words, $\partial V_{y_i}/\partial K_{y_\alpha}$ can never decrease when K_{y_α} increases. This result can also be written as follows:

$$\frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \geq 0$$

If j^* is not unique, we have a situation where exactly the same expected profit is associated with two or more branches at the same node in the decision tree. This is shown in Fig. 4.2. The same expected profit occurs at branches B_α and B_β for some value of K_{y_α}, which we call $K^o_{y_\alpha}$. As we change K_{y_α} we will change our decision at the

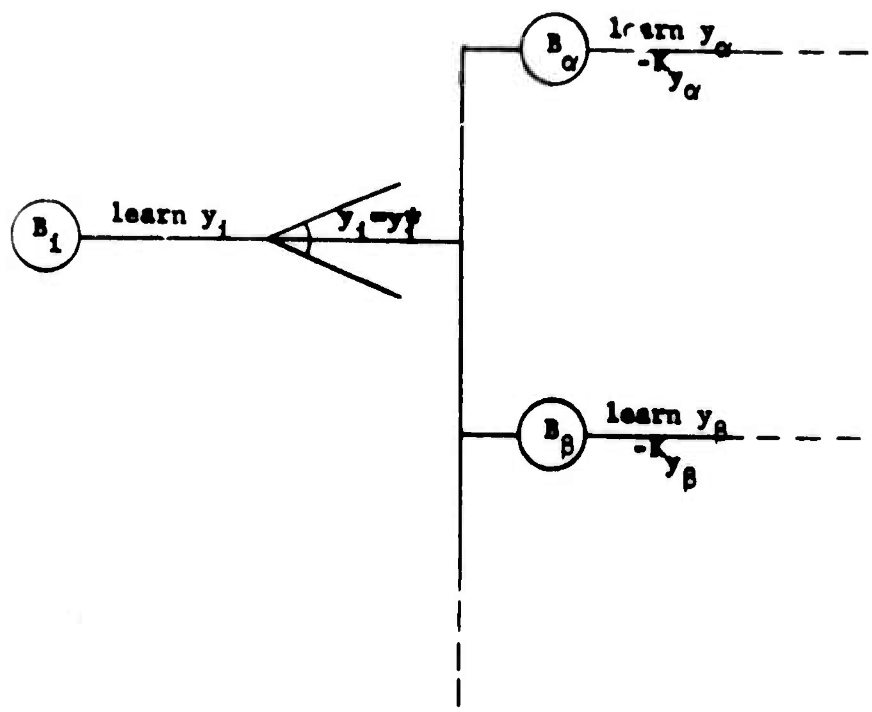


Figure 4.2a. Portion of a decision tree where the same expected profit occurs at two branches--B_α and B_β

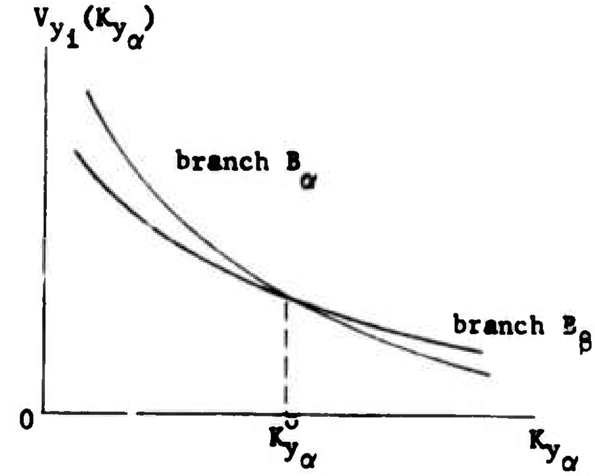


Figure 4.2b. Value of information associated with branch B_1 when branch B_α is always chosen, and when B_β is always chosen.

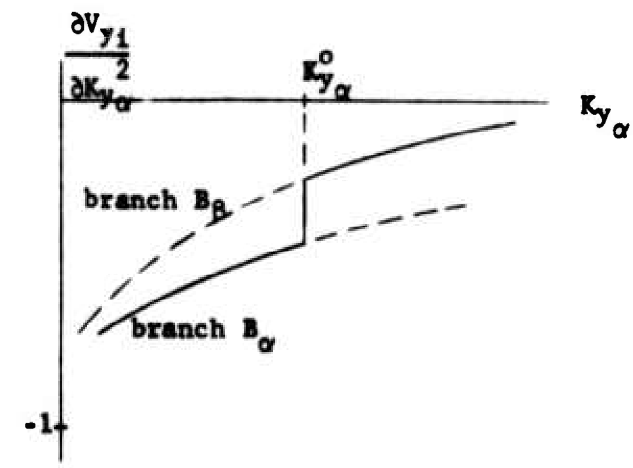


Figure 4.2c. First derivative of value of information associated with branch B_1 when branch B_α is always chosen, and when branch B_β is always chosen.

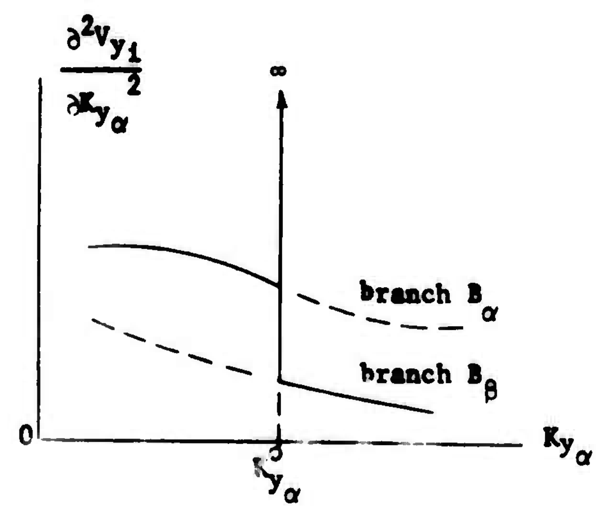


Figure 4.2d. Second derivative of value of information associated with branch B_1 when branch B_α is always chosen, and when branch B_β is always chosen.

node in question and choose a different branch. If we look at the expected profit and the corresponding value of information associated with branch B_i, assuming that branch B_α is always chosen over branch B_β, and vice versa, we must have a graph such as that shown in Fig. 2.4b. Since the proof above holds for any tree branches, both of the curves in Fig. 4.2b must have first derivatives in the interval $[-1,0]$, and positive second derivatives. This is shown in Fig. 4.2c and 4.2d. In order for the value of information curves in Fig. 4.2b to cross, the slope of the B_α curve must be less than the slope of the B_β curve at $K^o_{y_\alpha}$.

When we have a choice between branches B_α and B_β, we will have the value of information that is the envelope of the curves in Fig. 4.2b. This can result in a discontinuous first derivative, as shown by the solid curve in Fig. 4.2c. Since the slope of the B_α curve must be less than the slope of the B_β curve when K_{y_α} equals $K^o_{y_\alpha}$ in Fig. 4.2b, the first derivative must increase at the discontinuity. Therefore, the second derivative can contain a positive impulse, as shown by the solid curve in Fig. 4.2d. However, neither the discontinuity in $\partial V_{y_i}/\partial K_{y_\alpha}$ nor the impulse in $\partial^2 V_{y_i}/\partial K_{y_\alpha}{}^2$ invalidate the statement that

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \in [-1,0] \quad \text{and} \quad \frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \geq 0$$

We can reach the same conclusion algebraically, but the method of proof is exactly the same as the graphical argument given above. To avoid unnecessary complication, we will not go through the algebraic proof. (Using an algebraic proof we can show that $\partial V_{y_i}/\partial K_{y_\alpha}$ can make

a discontinuous jump to lower value as K_{y_β} increases, Thus, $\partial^2 V_{y_i}/\partial K_{y_\alpha}\partial K_{y_\beta}$ can contain a finite number of negative impulses, but it is non-negative elsewhere.)

In our proof that the first and second derivatives of V_{y_i} had certain properties, we assumed that j^* was equal to α. If j^* does not equal α we will still get the same conditions for $\partial V_{y_i}/\partial K_{y_\alpha}$ and $\partial^2 V_{y_i}/\partial K_{y_\alpha}{}^2$, but it will require some more effort. In this case,

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \underset{y_i}{E}\left\{\begin{array}{l} 0 : y_i \in \varphi_1(K_{y_{j*}}) \\ \underset{y_{j*}}{E}\left\{\begin{array}{l} 0 : \underset{y_i}{D}\,\underset{y_{j*}}{D}\,\max_c \underset{x_1}{E}\cdots\underset{x_n}{E}\,\pi \geq \underset{y_i}{D}\,\underset{y_{j*}}{D}\,\max_{k\neq i,j*}(\underset{y_k}{E}\,\max\{\ldots\}-K_{y_k}) \\ \frac{\partial}{\partial K_{y_\alpha}}\max_{k\neq i,j*}(\underset{y_k}{E}\,\max\{\ldots\}-K_{y_k}) : \text{otherwise} \end{array}\right\} \end{array}\right\}$$

(To save space, the condition "otherwise" is dropped when the meaning of the expression is clear.) Let k^* be the value of k that maximizes

$$\underset{y_i}{D}\,\underset{y_{j*}}{D}\left(\underset{y_k}{E}\,\max\left\{\begin{array}{l}\max_c \underset{x_1}{E}\cdots\underset{x_n}{E}\,\pi \\ \max_{\ell\neq i,j*,k}(\underset{y_k}{E}\,\max\{\ldots\}-K_{y_\ell})\end{array}\right\}-K_{y_k}\right)$$

and assume for a moment that k^* is unique. Thus the derivative becomes

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \underset{y_i}{E}\left\{\begin{array}{l} 0 : y_i \in \varphi_1(K_{y_{j*}}) \\ \underset{y_{j*}}{E}\left\{\begin{array}{l} 0 : \underset{y_i}{D}\,\underset{y_{j*}}{D}\,\max_c \underset{x_1}{E}\cdots\underset{x_n}{E}\,\pi \geq \underset{y_i}{D}\,\underset{y_{j*}}{D}\,\underset{y_{k*}}{E}\,\max\{\ldots\}-K_{y_{k*}} \\ \frac{\partial}{\partial K_{y_\alpha}}\underset{y_{k*}}{E}\,\max\{\ldots\}-K_{y_{k*}} : \text{otherwise} \end{array}\right\} \end{array}\right\}$$

If $k^* = \alpha$, we have

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \mathop{E}_{y_i} \left\{ \begin{array}{l} 0 : y_i \in \varphi_1(K_{y_{j*}}) \\ \mathop{E}_{y_{j*}} \left\{ \begin{array}{l} 0 : \mathop{D}_{y_i} \mathop{D}_{y_{j*}} \max_c \mathop{E}_{x_1} \cdots \mathop{E}_{x_n} \pi \geq \mathop{D}_{y_i} \mathop{D}_{y_{j*}} \mathop{E}_{y_\alpha} \max\{\ldots\} - K_{y_\alpha} \\ -1 : \text{otherwise} \end{array} \right\} \end{array} \right\}$$

Define $\theta(K_{y_\alpha})$ as follows

$$\theta(K_{y_\alpha}) = \mathop{D}_{y_i} \mathop{E}_{y_{j*}} \left\{ \begin{array}{l} 0 : \mathop{D}_{y_i} \mathop{D}_{y_{j*}} \max_c \mathop{E}_{x_1} \cdots \mathop{E}_{x_n} \pi \geq \mathop{D}_{y_i} \mathop{D}_{y_{j*}} \mathop{E}_{y_\alpha} \max\{\ldots\} - K_{y_\alpha} \\ -1 : \text{otherwise} \end{array} \right\}$$

The expected value of a quantity that is everywhere either 0 or -1 must lie in the closed interval bounded by these numbers. Therefore,

$$\theta(K_{y_\alpha}) \in [-1,0]$$

The derivative of V_{y_i} can be written as follows:

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \mathop{E}_{y_i} \left\{ \begin{array}{l} 0 : y_i \in \varphi_1(K_{y_{j*}}) \\ \theta(K_{y_\alpha}) : \text{otherwise} \end{array} \right\}$$

The expected value of a quantity that is everywhere in the interval [-1,0] must lie in this interval. Therefore,

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \in [-1,0]$$

Now consider what happens to $\partial V_{y_i}/\partial K_{y_\alpha}$ as K_{y_α} increases. Let $\varphi_2(K_{y_\alpha})$ be the set of all (y_i, y_{j*}) such that the condition in the equation for $\theta(K_{y_\alpha})$ is met.

$$\varphi_2(K_{y_\alpha}) = \Bigg[(y_i, y_{j*}) \; : \; K_{y_\alpha} \geq \underset{y_i}{D}\,\underset{y_{j*}}{D}\,\underset{y_\alpha}{E} \max \begin{Bmatrix} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} \pi \\ \max\limits_{\ell \neq i, j*, \alpha} (E \max\{\ldots\} - K_{y_\ell}) \end{Bmatrix} - \underset{y_i}{D}\,\underset{y_{j*}}{D} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} \pi \Bigg]$$

We can rewrite the equation for $\theta(K_{y_\alpha})$ as follows:

$$\theta(K_{y_\alpha}) = \underset{y_i}{D}\,\underset{y_{j*}}{E} \begin{Bmatrix} 0 & : (y_i, y_{j*}) \in \varphi_2(K_{y_\alpha}) \\ -1 & : \text{otherwise} \end{Bmatrix}$$

Assume that (y_i^o, y_{j*}^o) belongs to $\varphi_2(K_{y_\alpha})$, and let K'_{y_α} be some price larger than K_{y_α}. Then,

$$K'_{y_\alpha} > K_{y_\alpha} \geq \underset{y_i^o}{D}\,\underset{y_{j*}^o}{D}\,\underset{y_\alpha}{E} \max \begin{Bmatrix} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} \pi \\ \max\limits_{\ell \neq i, j*, \alpha} (\underset{y_\ell}{E} \max\{\ldots\} - K_{y_\ell}) \end{Bmatrix}$$

$$- \underset{y_i^o}{D}\,\underset{y_{j*}^o}{D} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} \pi$$

Therefore (y_i^o, y_{j*}^o) belongs to $\varphi_2(K'_{y_\alpha})$ as well as to $\varphi_2(K_{y_\alpha})$. Since this is true for any (y_i^o, y_j^o) that belongs to $\varphi_2(K_{y_\alpha})$

$$\varphi_2(K_{y_\alpha}) \in \varphi_2(K'_{y_\alpha})$$

whenever $K'_{y_\alpha} > K_{y_\alpha}$. So,

$$\theta(K'_{y_\alpha}) = \underset{y_i}{D}\ \underset{y_{j*}}{E} \left\{ \begin{array}{ll} 0 & : (y_i, y_{j*}) \in \varphi_2(K'_{y_\alpha}) \\ -1 & : \text{otherwise} \end{array} \right\}$$

$$\geq \underset{y_i}{D}\ \underset{y_{j*}}{E} \left\{ \begin{array}{ll} 0 & : (y_i, y_{j*}) \in \varphi_2(K_{y_\alpha}) \subseteq \varphi_2(K'_{y_\alpha}) \\ -1 & : \text{otherwise} \end{array} \right\} = \theta(K_{y_\alpha})$$

In other words, $\theta(K_{y_\alpha})$ cannot decrease when K_{y_α} increases. Therefore,

$$\frac{\partial V_{y_i}}{\partial K'_{y_\alpha}} = \underset{y_i}{E} \left\{ \begin{array}{ll} 0 : y_i \in \varphi_1(K_{y_{j*}}) \\ \theta(K'_{y_\alpha}) : \text{otherwise} \end{array} \right\} \geq \underset{y_i}{E} \left\{ \begin{array}{ll} 0 : y_i \in \varphi_1(K_{y_{j*}}) \\ \theta(K_{y_\alpha}) : \text{otherwise} \end{array} \right\}$$

$$= \frac{\partial V_{y_i}}{\partial K_{y_\alpha}}$$

This follows from the fact that the expected value of a quantity cannot decrease when the quantity is increased or held constant. Thus $\partial V_{y_i}/\partial K_{y_\alpha}$ cannot decrease when K_{y_α} increases, and we can write

$$\frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \geq 0$$

If k^* is not unique, we have a situation where exactly the same expected profit is associated with two or more branches at the same node of the decision tree. The same argument that we used previously shows

that $\partial V_{y_i}/\partial K_{y_\alpha}$ can have a discontinuity, and that $\partial^2 V_{y_i}/\partial K_{y_\alpha}^2$ can have a positive impulse at the value of K_{y_α} where the expected profits of the branches are equal. However, it is still true that

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \in [-1,0] \quad \text{and} \quad \frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \geq 0$$

If k^* does not equal α, we repeat the procedure, adding new observables until we find one that maximizes the appropriate expression and is equal to y_α. This must occur within n iterations since each observable we add must be different from all of the preceding observables. When we finally get an expression for $\partial V_{y_i}/\partial K_{y_\alpha}$ that contains K_{y_α}, it will look like this :

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = \underset{y_i}{E}\left\{\begin{array}{l} 0 : y_i \in \varphi_1(K_{y_{j^*}}) \\ \underset{y_{j^*}}{E}\left\{\begin{array}{l} 0 : (y_i, y_{j^*}) \in \varphi_2(K_{y_{k^*}}) \\ \quad 0 : (y_i, y_{j^*}, y_{k^*}) \in \varphi_3(K_{y_{\ell^*}}) \\ \underset{y_{k^*}}{E}\left\{ \cdots \underset{y_\alpha}{E}\left\{\begin{array}{l} 0 : (y_i, y_{j^*}, \ldots) \in \varphi_r(K_{y_\alpha}) \\ -1 : \text{otherwise} \end{array}\right\}\right\} \end{array}\right\} \end{array}\right\}$$

Following the procedures used previously we can show that the innermost expectation must lie in the interval [-1,0], and that it will not decrease when K_{y_α} increases. Then we can work back through the other expected values to show that

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \in [-1,0] \quad \text{and} \quad \frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \geq 0$$

The Number of Calculations Required to Determine the Value Functions and Decision Regions

Unless a sequential information problem is very simple, the number of calculations required to determine the decision regions is prohibitively large. Furthermore, when there are more than three observables it becomes very difficult to even visualize the decision regions because they occupy a space of more than three dimensions.

How many calculations are required to determine the decision regions in a problem with n observables? To answer this question, consider a problem with three observables. Following the procedure developed in Chapters 2 and 3, we first analyze the decision of whether or not to buy information about one observable when we already know the other two. The solution to this problem is a function of the two observables that we already know. There are three subordinate decision problems of this type corresponding to the three observables. Figure 4.3 shows a simplified decision tree for this problem with branches that lead to the same state of information connected togeter. (It is a feature of decision problems that different sets of decisions can lead to the same state of information. This feature is called "coalescence." In terms of decision trees, coalescence means that different branches can be connected since they are identical beyond a certain point.) The nodes at which we must solve a subordinate, information-purchasing decision problem are surrounded by small rectangles. The three subordinate decision problems discussed above are Nodes 5, 6, and 7 in Fig. 4.3.

Next we back up one step and consider the information-purchasing decisions at those nodes where we know one of the observables. At these

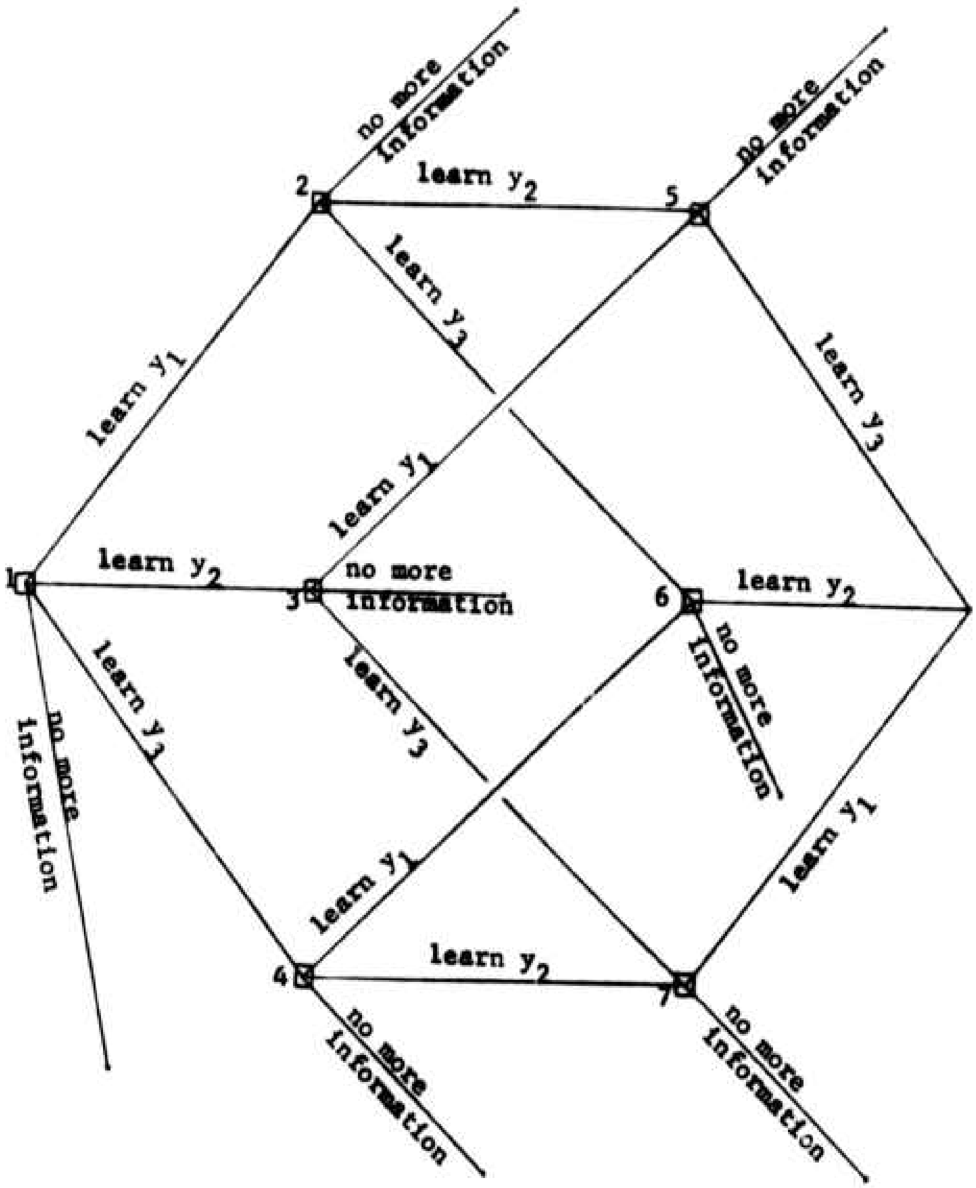


Figure 4.3. Simplified decision tree for a problem with three observables

nodes we are trying to decide which, if any, of the remaining two observables to buy. The expected profit associated with learning another observable is given by the solutions of the previous set of subordinate decision problems, the ones where we assumed that we knew two of the observables. There are three subordinate decision problems where we only know one of the observables, and the solution to each of these problems is a function of the one observable that is known. These three subordinate decision problems correspond to Nodes 2, 3, and 4 in Fig. 4.3.

Finally we have to back up to the first node in the decision tree (Node 1 in Fig. 4.3) and decide which piece of information, if any, to buy first. The expected profit associated with learning any observable is given by the solutions to the previous subordinate decision problems.

When we solve all of the subordinate decision problems associated with the nodes in Fig. 4.3, we will have the inequalities that describe the decision regions in n-dimensional space. If we solve all of the subordinate decision problems except the first one (Node 1 in Fig. 4.3), take the appropriate expected values, and then difference the results, we will have $V_{y_1}(K_{y_2},K_{y_3},K_{y_4})$ for $i = 1,2,3$. For the problem with three observables, this means that we need to solve six subordinate decision problems to find the value of sequential information about each of the observables. We must solve one more decision problem to determine the decision regions.

If we have a decision problem with n observables, we will have to solve n subordinate decision problems where we assume that we know all but one of the observables. This is followed by $n(n-1)/2$ subordinate decision problems where we assume that we know all but two of

the observables. In general, if we assume that we know k observables, we will have

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

subordinate decision problems to solve. This is the number of combinations of n distinct things that we can take k at a time. Since we must solve these decision problems for $k = 2,3,\ldots,(n-1)$ to find $V_{y_i}(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n})$, we will have to solve a total of

$$\sum_{k=2}^{n-1} \binom{n}{k} = (2^n - 2)$$

subordinate decision problems. To determine the decision regions (or decision sets Ω_{y_i}) we will need to solve one more subordinate decision problem for a total of $(2^n - 1)$ problems.

Approximating and Characterizing the Decision Regions

As we have seen, a large number of calculations are required to determine the value functions and the associated decision regions when there are many observables. Since the value of sequential information and the boundaries of the decision regions are described by algebraic functions, the calculations required to determine them are algebraic rather than numeric. Therefore it is difficult to program a computer to carry out the calculations.

However, we can use the decision tree in Fig. 4.1 to determine $V_{y_i}(K_{y_1},\ldots,K_{y_{i-1}},K_{y_{i+1}},\ldots,K_{y_n})$, for $i = 1,\ldots,n$, and the best initial information-purchasing decision whenever we have a specific set of

prices $(K_{y_1},\ldots,K_{y_n})$. The calculations required to solve the decision tree are numeric, and a computer program can be written to solve the decision tree. Any computer program that solves the decision tree for a sequential information problem with n observables must solve $(2^n - 1)$ subordinate decision problems to determine the value of information and the best initial information-purchasing decision. However, each subordinate decision problem can be solved by simply comparing numbers and choosing the largest one. This corresponds to picking the branch with the highest expected profit at each node of the simplified decision tree in Fig. 4.3.

By solving the decision tree in Fig. 4.1 with a specific set of prices for the observables, we are determining the optimum decision at one point in the price diagram. By repeating the calculation for a number of different sets of prices, we can determine the approximate boundaries of the decision regions. The values of individual and simultaneous information can be used to determine the sets of prices that would be most useful in determining the decision regions. For example, if we had determined $V^N_{y_1}$, $V^N_{y_2}$, and $V^N_{y_1 y_2}$ for a problem with two observables, as shown in Fig. 4.4, we would want to solve the decision tree with sets of prices represented by points in the region R to see if we would be willing to buy information in that region.

The values of individual and simultaneous information can be determined numerically using the decision tree in Fig. 4.1 and an appropriate computer program. The expected profit when no information is purchased can be found by solving the decision tree with all of the prices of the observables set equal to infinity (or some very large number). $V^N_{y_i}$ is

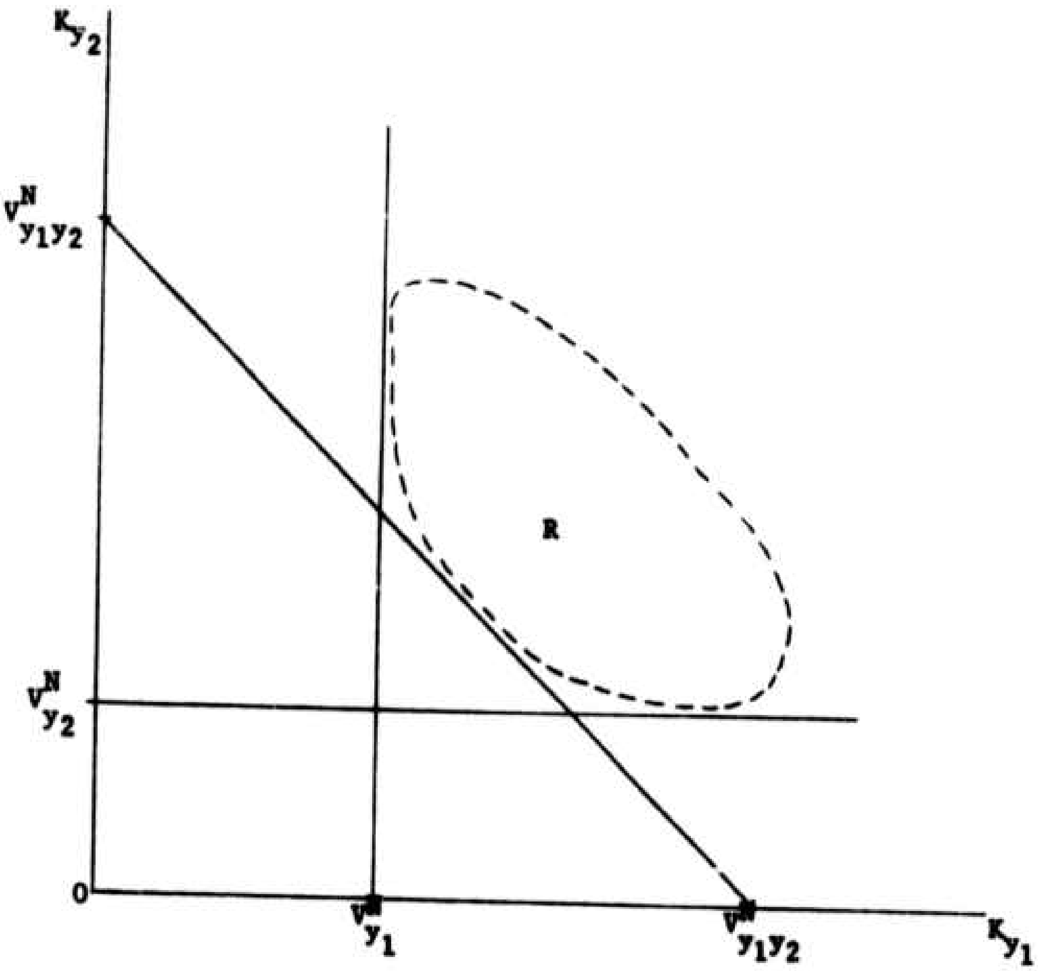


Figure 4.4. Price diagram showing region where it may be desirable to buy information sequentially

equal to the increase in expected profit that results from setting K_{y_i} equal to zero and all of the other prices equal to infinity. $V^N_{y_1 \ldots y_n}$ is the increase in expected profit that results from setting all of the prices equal to zero.

Another way to use the decision tree to characterize the decision regions is by bounding the maximum value that $V_{y_i}(K_{y_1}, \ldots, K_{y_n})$ could have when our best initial decision is to learn y_i. This maximum value can be used as an upper bound for the amount that we would be willing to pay to learn y_i first, regardless of what the prices of the other observables are. The upper bound allows us to exclude from consideration any initial purchase of information that costs more than the bound, without considering the price of each observable.

It is easy to show that the value of simultaneous information about all of the observables is an upper bound for the maximum value of sequential information about any one of the observables. When the prices of the observables are all set equal to zero, V_{y_i} equals $V^N_{y_1 \ldots y_n}$. This can be seen from the decision tree in Fig. 4.1. Since all of the information is free, we will continue to accept one piece of information after another until we have received it all, regardless of which observable we learn first. In this case the expected profit associated with learning all of the observables simultaneously equals the expected profit associated with learning them sequentially in any order. Therefore,

$$V_{y_i}(0,0,\ldots,0) = V^N_{y_1 \ldots y_n} \qquad (i = 1,\ldots,n)$$

As we increase each of the prices to any positive quantity, V_{y_i} cannot

increase since $\partial V_{y_i}/\partial K_{y_j}$ is less than or equal to zero. Thus,

$$V_{y_i}(K_{y_1},\ldots,K_{y_n}) \leq V^N_{y_1\ldots y_n}$$

Since $V^N_{y_1\ldots y_n}$ is an upper bound for V_{y_i} for all sets of positive prices, it must also be an upper bound for V_{y_i} over the set of prices where our best initial decision is to learn y_i.

Unfortunately, $V^N_{y_1\ldots y_n}$ is not a very tight upper bound for the maximum value that V_{y_i} can have when our best initial decision is to learn y_i. To find this maximum value, we have to maximize K_{y_i} subject to the constraint that the increase in expected profit resulting from learning y_i must be positive and must exceed the increase in expected profit resulting from learning any other observable. When there are more than two observables this is a very difficult optimization problem.

In addition, when there are three or more observables it is possible that the set of prices that maximizes K_{y_i} subject to

$$(V_{y_i} - K_{y_i}) \geq (V_{y_k} - K_{y_k})$$

for all k not equal to i, and

$$(V_{y_i} - K_{y_i}) \geq 0$$

is not the same as the set of prices that maximizes K_{y_j} subject to

$$(V_{y_j} - K_{y_j}) \geq (V_{y_k} - K_{y_k})$$

for all k not equal to j, and

$$(V_{y_j} - K_{y_j}) \geq 0$$

In other words, there need not be a unique point in the n-dimensional price diagram where the price of each observable is maximized over the decision region corresponding to that observable, even when there are no sets of prices such that the best initial decision is to buy several pieces of information simultaneously.

(If the best initial decision can be to buy several pieces of information simultaneously, there will be regions in the price diagram where $(V_{y_i} - K_{y_i})$ equals $(V_{y_j} - K_{y_j})$, and many sets of prices may maximize K_{y_i} and also meet the constraints. When buying simultaneous information cannot be the best initial decision, a unique set of prices that maximizes all of the prices over their corresponding decision regions must satisfy

$$V_{y_i} - K_{y_i} = 0 \quad \text{for} \quad i = 1,\ldots,n$$

Also, $\partial K_{y_i}/\partial K_{y_j}$ must be positive along the boundary where the decision regions meet, at least in the vicinity of the point at which all of the prices are maximized. However a simple calculation will show that this condition is only met when n equals two (see the following discussion). When n exceeds two, $\partial K_{y_i}/\partial K_{y_j}$ can be negative along the boundary where the decision regions meet, permitting the existence of different sets of prices which maximize the various K_{y_i} over their corresponding decision regions.)

When there are only two observables, as there were in the examples in Chapters 2 and 3, there is a unique pair of prices, $(K^*_{y_1}, K^*_{y_2})$, that maximizes the price of each observable over the corresponding decision region, provided that there are no pairs of prices such that the best initial decision is to buy both pieces of information simultaneously. (An example of a decision problem where buying both observables simultaneously can be the best initial decision is "Decision Problem One" in Chapter 1. In such problems it does not matter which observable we buy first because we will always

buy the other one.) In Fig. 4.5, the pair of prices $(K^*_{y_1}, K^*_{y_2})$ is represented by the point A. To see that the coordinates of the point A are the maximum values that K_{y_1} and K_{y_2} can have in their respective decision regions, consider the boundary A-B separating these regions. The equation for this boundary is

$$V_{y_1}(K_{y_2}) - K_{y_1} = V_{y_2}(K_{y_1}) - K_{y_2}$$

Differentiating this equation with respect to K_{y_1} yields

$$\left[\frac{\partial V_{y_1}}{\partial K_{y_2}}\right]\left[\frac{\partial K_{y_2}}{\partial K_{y_1}}\right] - 1 = \left[\frac{\partial V_{y_2}}{\partial K_{y_1}}\right] - \left[\frac{\partial K_{y_2}}{\partial K_{y_1}}\right]$$

Therefore,

$$\frac{\partial K_{y_2}}{\partial K_{y_1}} = \frac{(1 + \partial V_{y_2}/\partial K_{y_1})}{(1 + \partial V_{y_1}/\partial K_{y_2}}$$

Since $\partial V_{y_2}/\partial K_{y_1}$ and $\partial V_{y_1}/\partial K_{y_2}$ must lie in the interval $[-1,0]$, $\partial K_{y_2}/\partial K_{y_1}$ is equal to the ratio of two non-negative numbers. Thus $\partial K_{y_2}/\partial K_{y_1}$ cannot be negative along the boundary A-B in Fig. 4.5. Using this result and the fact that the expected value functions (which define the other boundaries in Fig. 4.5) decline as the prices increase, we can see that the maximum values of K_{y_1} and K_{y_2} in their respective decision regions must be the coordinates of the point A in Fig. 4.5. Thus for sequential information problems with two observables, we can characterize the decision regions by finding or bounding the coordinates of a single point in the price diagram.

The coordinates of point A in Fig. 4.5 are given by the solution

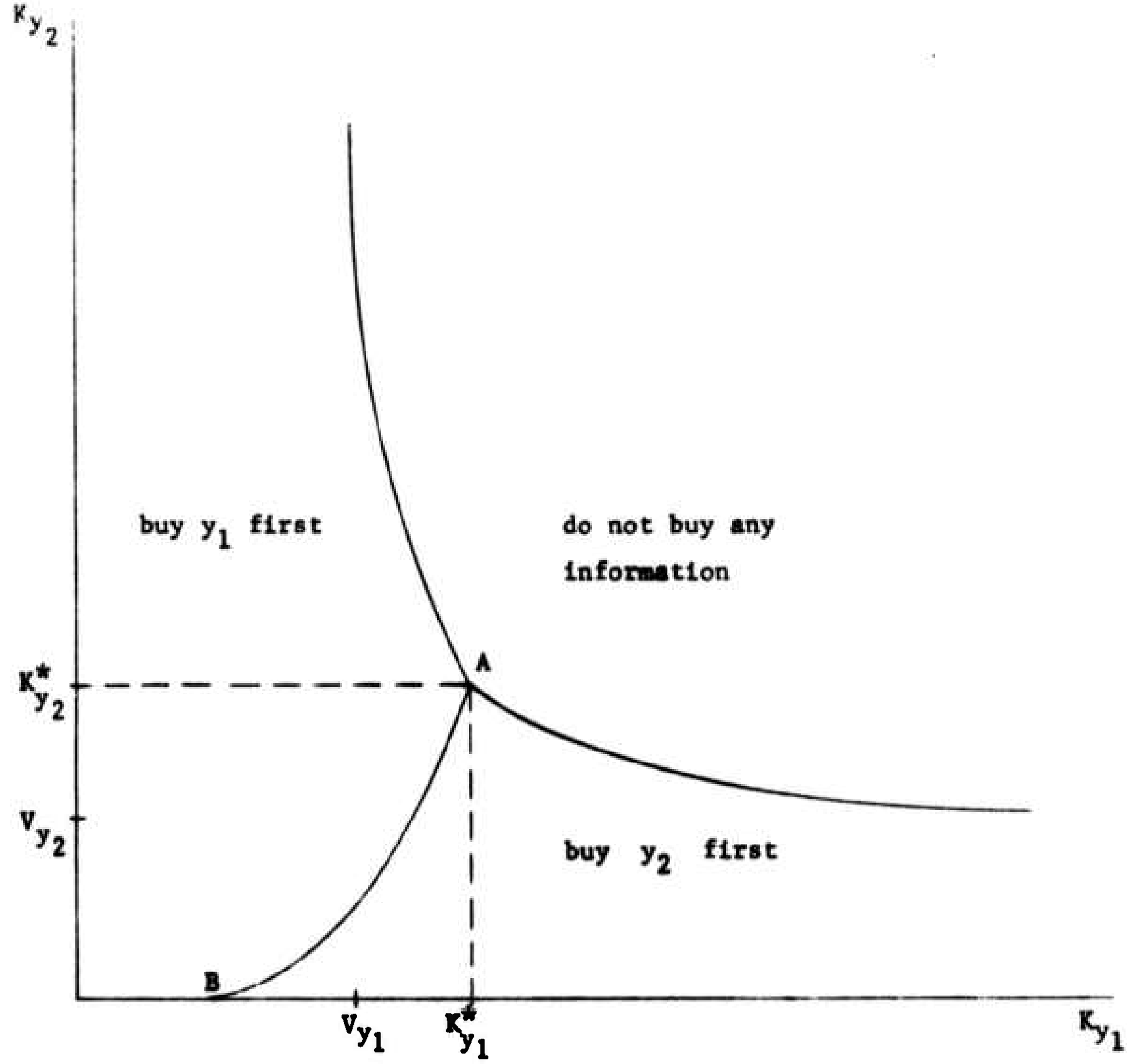


Figure 4.5. Price diagram showing maximum initial prices

of the following pair of equations

$$K^*_{y_1} = V_{y_1}(K^*_{y_2}) \quad \text{and} \quad K^*_{y_2} = V_{y_2}(K^*_{y_1})$$

Suppose we have a pair of prices (K'_{y_1}, K'_{y_2}) such that

$$K'_{y_1} \leq K^*_{y_1} \quad \text{and} \quad K'_{y_2} \leq K^*_{y_2}$$

We define another pair of prices as follows:

$$K''_{y_1} = V_{y_1}(K'_{y_2}) \quad \text{and} \quad K''_{y_2} = V_{y_2}(K'_{y_1})$$

Since $\partial V_{y_i}/\partial K_{y_j}$ cannot be positive, decreasing K_{y_j} from $K^*_{y_j}$ to K'_{y_j} cannot decrease V_{y_i}. Therefore,

$$K^*_{y_1} = V_{y_1}(K^*_{y_2}) \leq V_{y_1}(K'_{y_2}) = K''_{y_1}$$

$$K^*_{y_2} = V_{y_2}(K^*_{y_1}) \leq V_{y_2}(K'_{y_1}) = K''_{y_2}$$

Thus if we start with a pair of prices such that each is less than or equal to the corresponding coordinate of the point A in Fig. 4.5, we can find a pair of prices that are greater than or equal to the coordinates of the point A. In other words, if we start with a pair of lower bounds for the maximum prices of the observables over their decision regions, we can find a pair of upper bounds for these maximum prices.

We still are faced with the problem of finding a pair of prices (K'_{y_1}, K'_{y_2}) that are each less than or equal to the corresponding coordinate of the point A in Fig. 4.5. However, a pair of prices that meets

this requirement is easy to find since we have shown that $V_{y_i}^N$ is less than or equal to V_{y_i} for any pair of prices, and that $K^*_{y_i} = V_{y_i}(K^*_{y_j})$. Thus, if we let

$$K'_{y_1} = V_{y_1}^N \quad \text{and} \quad K'_{y_2} = V_{y_2}^N$$

we can be sure that K'_{y_i} is less than or equal to $K^*_{y_i}$ $(i = 1,2)$.

Using these values for (K'_{y_1}, K'_{y_2}) to get an upper bound for the maximum price of each observable over the corresponding decision region yields

$$K''_{y_1} = V_{y_1}(V_{y_2}^N) \geq K^*_{y_1}$$

$$K''_{y_2} = V_{y_2}(V_{y_1}^N) \geq K^*_{y_2}$$

Thus when there are two observables, we can solve the decision tree in Fig. 4.1b with the prices of the observables set equal to the values of individual information. This will give us upper bounds for the amount that we should pay initially for each observable when the information is available sequentially.

We can carry this procedure further, and develop an iterative procedure for determining $(K^*_{y_1}, K^*_{y_2})$ to any degree of accuracy. Define another set of prices as

$$K'''_{y_1} = V_{y_1}(K''_{y_2}) \quad \text{and} \quad K'''_{y_2} = V_{y_2}(K''_{y_1})$$

We can continue in this manner determining new pairs of prices, each of which is equal to V_{y_1} and V_{y_2} evaluated at the preceding pair of

prices. By repeating this procedure indefinitely, substituting each pair of prices into the decision tree in Fig. 4.1b to find a new pair of prices, we will determine a sequence of price pairs that converge to $(K^*_{y_1}, K^*_{y_2})$.

We can show that convergence will occur by using the properties of V_{y_i} that we derived previously. We know that

$$-1 \leq \frac{V_{y_1}(K_{y_2})}{K_{y_2}} \leq 0$$

$$-1 \leq \frac{V_{y_1}(K_{y_2})}{K_{y_2}} \leq 0$$

Since we have assumed that our best initial decision cannot be to buy both observables simultaneously, and hence that $(K^*_{y_1}, K^*_{y_2})$ must be unique, the two functions $V_{y_1}(\cdot)$ and $V_{y_2}(\cdot)$ cannot both have a slope of minus one in the vicinity of $(K^*_{y_1}, K^*_{y_2})$. If they did, the equations

$$K_{y_1} = V_{y_1}(K_{y_2}) \quad \text{and} \quad K_{y_2} = V_{y_2}(K_{y_1})$$

would have a solution at $(K^*_{y_1} + \epsilon \, , \, K^*_{y_2} - \epsilon)$ as well as at $(K^*_{y_1}, K^*_{y_2})$. Thus in the vicinity of $(K^*_{y_1}, K^*_{y_2})$ the following inequalities must be true:

$$-1 < \frac{\partial V_{y_1}(K_{y_2})}{\partial K_{y_2}} \leq 0$$

$$-1 \leq \frac{\partial V_{y_1}(K_{y_2})}{\partial K_{y_2}} \leq 0$$

where we have designated the function whose slope strictly exceeds minus one as $V_{y_1}(K_{y_2})$.

Since the derivatives of V_{y_1} and V_{y_2} must satisfy the inequalities above, we have

$$-1 < \left[\frac{V_{y_1}(K^*_{y_2}) - V_{y_1}(K'_{y_2})}{K^*_{y_2} - K'_{y_2}} \right] \leq 0$$

$$-1 \leq \left[\frac{V_{y_2}(K^*_{y_1}) - V_{y_2}(K'_{y_1})}{K^*_{y_2} - K'_{y_2}} \right] \leq 0$$

Simplifying these inequalities yields

$$K^*_{y_2} - K'_{y_2} > K''_{y_1} - K^*_{y_1} \geq 0$$

$$K^*_{y_1} - K'_{y_1} \geq K''_{y_2} - K^*_{y_2} \geq 0$$

Using the inequalities for the derivatives again,

$$-1 < \left[\frac{V_{y_1}(K''_{y_2}) - V_{y_1}(K^*_{y_2})}{K''_{y_2} - K^*_{y_2}} \right] \leq 0$$

$$-1 \leq \left[\frac{V_{y_2}(K''_{y_1}) - V_{y_2}(K^*_{y_1})}{K''_{y_1} - K^*_{y_1}} \right] \leq 0$$

Simplifying these inequalities yields

$$K''_{y_2} - K^*_{y_2} > K^*_{y_1} - K'''_{y_1} \geq 0$$

$$K''_{y_1} - K^*_{y_1} \geq K^*_{y_2} - K'''_{y_2} \geq 0$$

Now combine these with the preceding inequalities:

$$K^*_{y_1} - K'_{y_1} > K^*_{y_1} - K'''_{y_1} \geq 0$$

$$K^*_{y_2} - K'_{y_2} > K^*_{y_2} - K'''_{y_2} \geq 0$$

This can also be written as follows,

$$K'_{y_1} < K'''_{y_1} \leq K^*_{y_1}$$

$$K'_{y_2} < K'''_{y_2} \leq K^*_{y_2}$$

Thus after two iterations we have a pair of prices (K'''_{y_1}, K'''_{y_2}) that must lie closer to $(K^*_{y_1}, K^*_{y_2})$ than the pair of prices with which we started. If we continue the process we will find other pairs of prices, each of which is closer to the answer than the preceding pair. After enough iterations we will be arbitrarily close to the maximum price of each observable over its corresponding decision region.

We have been able to find a number of different ways to approximmate or characterize the decision regions. However, why should we bother to determine the decision regions, approximately or otherwise? When we are faced with an actual decision problem we will know the prices of the observables, or at least we will be able to encode our uncertainty about the prices. (The next chapter deals with the question of what to do when the prices of the observables are uncertain.) If we know the prices we will only be interested in that point in the price diagram that corresponds to them. We can determine our best initial information-purchasing decision by solving the decision tree in Fig. 4.1. We cannot rely on numbers like the values of individual and simultaneous information. We

do not really need to determine the decision regions for any practical decision problem. However, we do have to realize that they exist, and that the value of sequential information depends on the prices of the observables.

Utility and Risk Preference

Up to this point we have assumed that our objective is to maximize our expected profit. Now we will show that the conclusions of this chapter also apply to a decision maker whose utility function satisfies the delta property and is monotonically increasing.

Introducing a utility function into a sequential information decision problem does not alter the fact that, in general, V_{y_i} is a function of the prices of all of the observables except y_i. This dependence can be seen in the complete decision tree for the problem, which is shown in Fig. 4.6 for an arbitrary utility function. V_{y_i} is equal to the value of K_{y_i} such that the expected utility associated with branch B_i (the branch where we first learn y_i) is equal to the expected utility associated with branch B_0 (the branch where we do not learn any of the observables). Algebraically V_{y_i} is equal to the value of K_{y_i} such that

$$\mathop{E}_{y_i} \max \left\{ \begin{array}{l} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} U(\pi - K_{y_i}) \\ \max\limits_{j \neq i} \left(\mathop{E}\limits_{y_j} \max \left\{ \begin{array}{l} \max\limits_{c} \mathop{E}\limits_{x_1} \cdots \mathop{E}\limits_{x_n} U(\pi - K_{y_i} - K_{y_j}) \\ \max\limits_{k \neq i,j} \left(\mathop{E}\limits_{y_k} \max \{\ldots\} \right) \end{array} \right\} \right) \end{array} \right\}$$

$$= \max_{c} \mathop{E}_{x_1} \cdots \mathop{E}_{x_n} U(\pi)$$

for $i = 1,\ldots,n$.

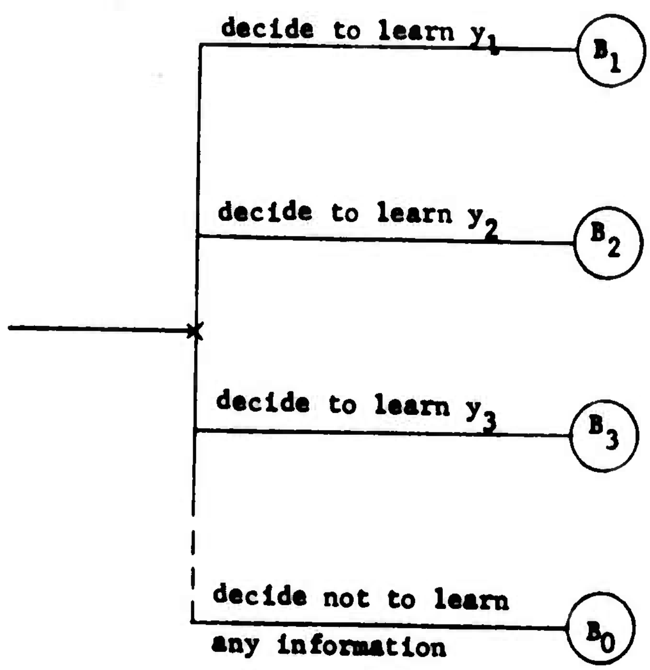


Figure 4.6a. Sequential information decision tree with arbitrary utility function

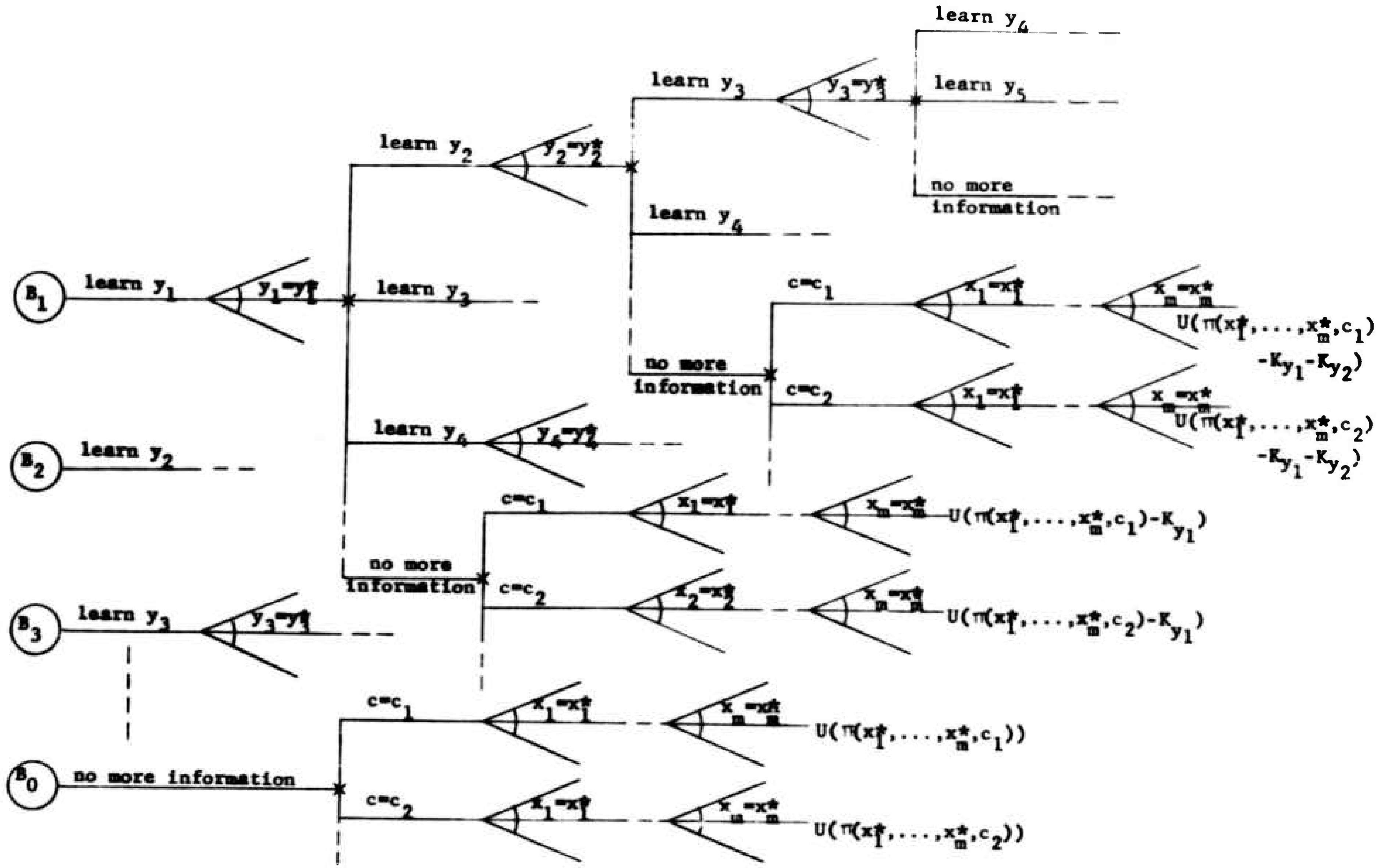


Figure 4.6b. Sequential information decision tree for arbitrary utility function (continued)

This relation holds for any utility function.

When the utility function satisfies the delta property and is monotonically increasing, we can solve the equation above for V_{y_i} ,

$$V_{y_i} = U^{-1}\left(\underset{y_i}{E} \max \left\{ \begin{array}{l} \max \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \\ \max\limits_{j \neq i} \left(\underset{y_j}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi - K_{y_j}) \\ \max\limits_{k \neq i,j} \left(\underset{y_k}{E} \max\{\dots\} \right) \end{array} \right\} \right) \end{array} \right\} \right)$$

$$- U^{-1}\left(\max_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)\right)$$

This equation can be rewritten in a form that separates the prices of the observables from the profit function.

$$V_{y_i} = U^{-1}\left(E_{y_i} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \\ \max\limits_{j \neq i} U \left[U^{-1} \left(\underset{y_j}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \\ \max\limits_{k \neq i,j} U[U^{-1}(\underset{y_k}{E} \max\{\dots\}) - K_{y_k}] \end{array} \right\} \right) - K_{y_j} \right] \end{array} \right\} \right)$$

$$- U^{-1}\left(\max_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)\right)$$

With the equation for V_{y_i} in this form, we can show that $\partial V_{y_i} / \partial K_{y_\alpha}$ must be negative or zero, and that $\partial^2 V_{y_i} / \partial K_{y_\alpha}{}^2$ must be positive or zero. The proof is similar to that given previously for an expected-profit decision maker, and for this reason it is only outlined here. Differentiating the equation above yields

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = (U^{-1})' \frac{\partial}{\partial K_{y_\alpha}} \left[\underset{y_i}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \\ \\ \max\limits_{j \neq i} U[U^{-1}(\underset{y_j}{E} \max\{\ldots\}) - K_{y_j}] \end{array} \right\} \right]$$

$$= (U^{-1})' \underset{y_i}{E} \left\{ \begin{array}{l} 0 : \underset{y_i}{D} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \geq \underset{y_i}{D} \max\limits_{j \neq i} U[U^{-1}(\underset{y_j}{E} \max\{\ldots\}) - K_{y_j}] \\ \\ \frac{\partial}{\partial K_{y_\alpha}} \max\limits_{j \neq i} U[U^{-1}(\underset{y_j}{E} \max\{\ldots\}) - K_{y_j}] : \text{otherwise} \end{array} \right\}$$

Let j^* be the value of j that maximizes

$$\underset{y_i}{D} U \left[U^{-1} \left(\underset{y_j}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \\ \\ \max\limits_{k \neq i,j} U[U^{-1}(\underset{y_k}{E} \max\{\ldots\}) - K_{y_k}] \end{array} \right\} \right) - K_{y_j} \right]$$

and assume for a moment that j^* is unique. In this case the derivative becomes

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = (U^{-1})' \underset{y_i}{E} \left\{ \begin{array}{l} 0 : \underset{y_i}{D} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) \geq \underset{y_i}{D} U[U^{-1}(\underset{y_{j*}}{E} \max\{\ldots\}) - K_{y_{j*}}] \\ \\ \frac{\partial}{\partial K_{y_\alpha}} U[U^{-1}(\underset{y_{j*}}{E} \max\{\ldots\}) - K_{y_{j*}}] : \text{otherwise} \end{array} \right\}$$

If j^* equals α,

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} = (U^{-1})' \underset{y_i}{E} \left\{ \begin{array}{l} 0 : \underset{y_i}{D} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi) = \underset{y_i}{D} U[U^{-1}(\underset{y}{E} \max\{\ldots\}) - K_{y_\alpha}] \\ \\ -U' : \text{otherwise} \end{array} \right\}$$

U' and $(U^{-1})'$ are both positive quantities, regardless of their arguments, because the utility function is monotonically increasing. The expected value of a quantity that is everywhere either zero or negative

cannot be positive. Therefore,

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \le 0$$

We can use the same argument that we used before to show that $\partial V_{y_i}/\partial K_{y_\alpha}$ cannot decrease when K_{y_α} increases, and therefore,

$$\frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \ge 0$$

Similarly, the same argument that we used previously shows that when j^* is not unique, or when there are two branches in the decision tree with the same expected profit, $\partial V_{y_i}/\partial K_{y_\alpha}$ can have a discontinuous increase and $\partial^2 V_{y_i}/\partial K_{y_\alpha}^2$ can contain a positive impulse. However, it is still true that

$$\frac{\partial V_{y_i}}{\partial K_{y_\alpha}} \le 0 \quad \text{and} \quad \frac{\partial^2 V_{y_i}}{\partial K_{y_\alpha}^2} \ge 0$$

If j^* does not equal α, we can get the same conditions for the first and second derivatives of V_{y_i} by extending the proof in exactly the same way that we did before.

We have proven that $\partial V_{y_i}/\partial K_{y_j}$ is less than or equal to zero, and it remains to be shown that this quantity must lie in the interval $[-1,0]$. We can prove that $\partial V_{y_i}/\partial K_{y_j} \ge -1$ by using implicit differentiation when there are only two observables, and we can use a qualitative argument to extend this result to problems with more than two observables.

When there are only two observables $V_{y_1}(K_{y_2})$ is given by the solution of the following equation:

$$\underset{y_1}{E} \max \begin{Bmatrix} U[U^{-1}(\max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)) - V_{y_1}(K_{y_2})] \\ U[U^{-1}(\underset{y_2}{E} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)) - V_{y_1}(K_{y_2}) - K_{y_2}] \end{Bmatrix}$$

$$= \max_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)$$

Differentiating both sides with respect to K_{y_2} yields

$$\frac{\partial}{\partial K_{y_2}} \underset{y_1}{E} \max \begin{Bmatrix} U[U^{-1}(\max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)) - V_{y_1}(K_{y_2})] \\ U[U^{-1}(\underset{y_2}{E} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)) - V_{y_1}(K_{y_2}) - K_{y_2}] \end{Bmatrix} = 0$$

Thus,

$$\underset{y_1}{E} \begin{Bmatrix} \frac{\partial}{\partial K_{y_2}} U[U^{-1}(\max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)) - V_{y_1}(K_{y_2})] : y_1 \in \varphi \\ \frac{\partial}{\partial K_{y_2}} U[U^{-1}(\underset{y_2}{E} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi)) - V_{y_1}(K_{y_2}) - K_{y_2}] : y_1 \notin \varphi \end{Bmatrix} = 0$$

where

$$\varphi = [\, y_1 : \underset{y_1}{D} \max \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi - V_{y_1}(K_{y_2}))$$

$$\geq \underset{y_1}{D} \underset{y_2}{E} \max_{c} \underset{x_1}{E} \cdots \underset{x_n}{E} U(\pi - V_{y_1}(K_{y_2}) - K_{y_2}) \,]$$

$$\underset{y_1}{E}\left\{\begin{array}{ll} U'[U^{-1}(\max_c \underset{x_1}{E} \cdots \underset{x_n}{E}\ U(\pi)) - V_{y_1}(K_{y_2})]\left[\dfrac{-V_{y_1}}{K_{y_2}}\right] & : y_1 \in \varphi \\ \\ U'[U^{-1}(\underset{y_2}{E}\ \max_c \underset{x_1}{E} \cdots \underset{x_n}{E}\ U(\pi)) - V_{y_1}(K_{y_2})]\left[\dfrac{-V_{y_1}}{K_{y_2}} - 1\right] & : y_1 \notin \varphi \end{array}\right\} = 0$$

Define

$$\beta_1(y_1) = \underset{y_1}{D}\ U'[U^{-1}(\max_c \underset{x_1}{E} \cdots \underset{x_n}{E}\ U(\pi)) - V_{y_1}(K_{y_2})] \geq 0$$

$$\beta_2(y_1) = \underset{y_1}{D}\ U'[U^{-1}(\underset{y_2}{E}\ \max_c \underset{x_1}{E} \cdots \underset{x_n}{E}\ U(\pi)) - V_{y_1}(K_{y_2}) - K_{y_2}] \geq 0$$

$$\alpha_1 = \sum_{y_1 \in \varphi} \beta_1(y_1)\ \{y_1|\delta\} \geq 0$$

$$\alpha_2 = \sum_{y_1 \notin \varphi} \beta_2(y_1)\ \{y_1|\delta\} \geq 0$$

Using these definitions, we have

$$\underset{y_1}{E}\left\{\begin{array}{ll} \beta_1(y_1)\left[\dfrac{-V_{y_1}}{K_{y_2}}\right] & : y_1 \in \varphi \\ \\ \beta_2(y_1)\left[\dfrac{-V_{y_1}}{K_{y_2}} - 1\right] & : y_1 \notin \varphi \end{array}\right\} = 0$$

$$\alpha_1 \left[\frac{-\partial V_{y_1}}{\partial K_{y_2}}\right] + \alpha_2\left[\frac{-\partial V_{y_1}}{\partial K_{y_2}} - 1\right] = 0$$

Therefore,

$$\frac{\partial V_{y_1}}{\partial K_{y_2}} = \left[\frac{-\alpha_2}{\alpha_1 + \alpha_2}\right]$$

Since α_1 and α_2 cannot both be equal to zero, we have

$$\frac{\partial V_{y_1}}{\partial K_{y_2}} \in [-1,0]$$

Similarly, we can show that

$$\frac{\partial V_{y_2}}{\partial K_{y_1}} \in [-1,0]$$

This proof demonstrates that $\partial V_{y_i}/\partial K_{y_j} \geq -1$ when there are only two observables. We will now show that this result must be true for any number of observables when the decision maker's utility function is monotonically increasing. Assume that there is a set of prices $(K^o_{y_1},\ldots,K^o_{y_n})$, such that the decision maker's preference is to buy y_i first, and $\partial V_{y_i}/\partial K_{y_j} < -1$. In this case we can decrease the price of y_i from $K^o_{y_i}$ to $(K^o_{y_i} - \epsilon)$, and the decision maker will increase the amount he is willing to pay for y_i from $V_{y_i}(K^o_{y_1},\ldots,K^o_{y_n})$ to $[V_{y_i}(K^o_{y_1},\ldots,K^o_{y_n}) + \delta]$, where $\delta > \epsilon$. If the prices of the observables are $(K^o_{y_1},\ldots,K^o_{y_n})$, a broker could make money by buying y_i for $K^o_{y_i}$ and offering to sell the observables to the decision maker for $(K^o_{y_1},\ldots,K^o_{y_i} + \delta,\ldots,K^o_{y_j} - \epsilon,\ldots,K^o_{y_n})$. The decision maker would be willing to buy y_i first for $(K^o_{y_i} + \delta)$, giving the broker a profit of δ. The broker would still be obligated to buy any additional observables at the original prices and sell them to the decision maker. However, the broker would lose at most ϵ by doing so. Thus he would make a profit of at least $(\delta - \epsilon)$.

A decision maker who finds himself in this situation can increase his assets, and hence his expected utility, by acting as his own broker.

Thus, whenever the prices of the observables are $(K^o_{y_1},\dots,K^o_{y_n})$ he is willing to pay $[V_y\ (K^o_{y_1},\dots,K^o_{y_n}) + \delta]$ to learn y_i first. However, this contradicts the fact that V_{y_i} is the maximum amount the decision maker will pay for y_i . Since our assumption that $\partial V_{y_i}/\partial K_{y_j} < -1$ has led to a contradiction, it cannot be valid.

Combining this result with the proof that $\partial V_{y_i}/\partial K_{y_j} \le 0$, we have

$$\frac{\partial V_{y_i}}{\partial K_{y_j}} \in [-1,0]$$

whenever the decision maker's utility function satisfies the delta property and is monotonically increasing.

It is clear that V_{y_i} must exceed $V^N_{y_i}$ and $V^R_{y_i}$, regardless of the form of the decision maker's utility function. The utility function does not alter the fact that we can achieve the expected utility associated with individual or simultaneous information by making the proper set of sequential decisions. Thus we must be willing to pay at least as much for sequential information about an observable as we would for individual or simultaneous information. (The amount we would pay for simultaneous information is the residual value of the information.)

It is also easy to see that introducing a utility function does not alter the number of subordinate decision problems that we must solve to find the value functions and the decision regions. However, it may be considerably more difficult to solve each subordinate decision problem when we have to take into account the decision maker's utility function, especially if his utility function does not satisfy the delta property.

Regardless of the form of the decision maker's utility function, we

can use the decision tree in Fig. 4.6 to determine V_{y_i} and the best initial information-purchasing decision for any given set of prices, $(K_{y_1},\ldots,K_{y_n})$. If the utility function does not satisfy the delta property, solving for the value of K_{y_i} that equalizes the expected utilities associated with branches B_i and B_0 in Fig. 4.6 will require a number of trial-and-error calculations. However this is unnecessary if the delta property applies. By solving the decision tree with a number of different sets of prices, we can determine the approximate boundaries of the decision regions in the price diagram.

When the decision maker's utility function satisfies the delta property and is monotonically increasing, we can use the fact that $\partial V_{y_i}/\partial K_{y_j}$ cannot be positive to show that

$$V_{y_i}(K_{y_1},\ldots,K_{y_n}) \le V^N_{y_1\ldots y_n}$$

Thus we can use the value of learning all of the observables simultaneously as an upper bound for the maximum value of learning y_i sequentially. However, $V^N_{y_1\ldots y_n}$ is still a weak upper bound for V_{y_i} over its associated decision region.

When there are only two observables, we can use the argument presented earlier to show that the maximum value of V_{y_i} over its associated decision region is bounded by $V_{y_i}(V^N_{y_j})$ since

$$\frac{\partial V_{y_i}}{\partial K_{y_j}} \in [-1,0] \qquad \left(\begin{array}{c} i,j = 1,2 \\ i\neq j \end{array}\right)$$

Thus we are able to extend the results of this chapter to the case

where the decision maker's utility function satisfies the delta property and is monotonically increasing.

Summary

In this chapter we have proven the following statements:

1. The value of sequential information is a function of the prices of the observables.
2. $V_{y_i} \geq V^N_{y_i}$ and $V_{y_i} \geq V^R_{y_i}$ for any set of prices $(K_{y_1}, \ldots K_{y_n})$.
3. V_{y_i} exceeds both $V^N_{y_i}$ and $V^R_{y_i}$ whenever learning y_i can affect our decision to buy another observable.
4. $\partial V_{y_i}/\partial K_{y_j} \in [-1,0]$ when the decision maker's utility function is monotonically increasing and satisfies the delta property.
5. $\partial^2 V_{y_i}/\partial K_{y_j}^2 \geq 0$ when the decision maker's utility function is monotonically increasing and satisfies the delta property.
6. If there are n observables, we must solve (2^n-1) subordinate decision problems to determine the value functions V_{y_i} and the decision regions.
7. $V^N_{y_1 \ldots y_n}$ is an upper bound for the value of sequential information.
8. When there are only two observables, V_{y_i} is bounded by $V_{y_i}(K_{y_j})$ for all pairs of prices where our best initial decision is to buy y_i. This holds when the decision maker's utility function is monotonically increasing and satisfies the delta property.

CHAPTER 5

SEQUENTIAL INFORMATION WITH UNCERTAIN AND NON-ADDITIVE PRICES

In this chapter we extend the results of the previous chapters to include problems where the cost of information is either uncertain or not additive. The two types of prices--uncertain and not additive--are considered separately. However we shall see that under certain conditions both cases can be analyzed using simple modifications of the procedures developed previously. We use the same notation in this chapter that we used before. Utility and risk aversion are not considered in this chapter.

The analysis of sequential information problems with uncertain and non-additive prices is first illustrated by considering simple modifications of the bidding problem introduced in Chapter 2. Then the results are generalized to apply to all decision problems.

The Bidding Problem with Uncertain Prices

Suppose that we face the bidding decision described in Chapter 2, but we are uncertain about the cost of learning our production cost p or the lowest competing bid ℓ. Thus, in addition to the probability density functions that describe our state of information about p and ℓ, we also have probability density functions for the prices K_p and K_ℓ. Assume for a moment that we have assessed p, ℓ, K_p, and K_ℓ to all be independent random variables. (We will relax this assumption later in this chapter.) Thus

$$\{p,\ell,K_p,K_\ell|\mathcal{E}\} = \{p|\mathcal{E}\}\ \{\ell|\mathcal{E}\}\ \{K_p|\mathcal{E}\}\ \{K_\ell|\mathcal{E}\}$$

As in Chapter 2, we assume that $\{p|\mathcal{E}\}$ is a uniform distribution between zero and one, and that $\{\ell|\mathcal{E}\}$ is a uniform distribution between zero and two. These distributions are shown in Fig. 2.1. We shall allow $\{K_p|\mathcal{E}\}$ and $\{K_\ell|\mathcal{E}\}$ to be arbitrary probability density functions.

If we do not receive any information about p and ℓ our expected profit i

$$\max_b \mathop{E}_p \mathop{E}_\ell \pi(p,\ell,b) = \max_b \mathop{E}_p \mathop{E}_\ell \begin{Bmatrix} b-p & : & b < \ell \\ 0 & : & b \geq \ell \end{Bmatrix} = 27/96$$

If we pay K_p for perfect information about p, and do not intend to learn ℓ, then our expected profit after learning p is

$$\max_b \mathop{E}_\ell \pi(p,\ell,b) - K_p$$

Before we know p and K_p our expected profit is

$$\mathop{E}_p \mathop{E}_{K_p} [\max_b \mathop{E}_\ell \pi(p,\ell,b) - K_p] = \mathop{E}_p \max_b \mathop{E}_\ell \pi(p,\ell,b) - \mathop{E}_{K_p} K_p$$

$$= \mathop{E}_p \max_b \mathop{E}_\ell \pi(p,\ell,b) - \overline{K}_p$$

We define $\overline{K}_p$ as follows

$$\overline{K}_p = \mathop{E}_{K_p} K_p = \langle K_p|\mathcal{E}\rangle$$

Therefore we should pay K_p to learn p by itself when

$$\overline{K}_p < V_p^N = (\mathop{E}_p \max_b \mathop{E}_\ell - \max_b \mathop{E}_p \mathop{E}_\ell)\ \pi(p,\ell,b) = 1/96$$

Using exactly the same reasoning we find that we should pay K_ℓ to learn ℓ by itself when

$$\overline{K}_\ell < V_\ell^N = (\mathop{E}_\ell \max_b \mathop{E}_p - \max_b \mathop{E}_p \mathop{E}_\ell)\ \pi(p,\ell,b) = 27/96$$

If we pay K_p plus K_ℓ to learn p and ℓ simultaneously, our expected profit after learning both state variables is

$$\max_b \pi(p,\ell,b) - K_p - K_\ell$$

(At this point we are still assuming that prices are additive. The case of non-additive prices will be considered later in this chapter.) Before we know the state variables or their prices, our expected profit is

$$\underset{p}{E}\,\underset{\ell}{E}\,\underset{K_p}{E}\,\underset{K_\ell}{E}\,[\max_b \pi(p,\ell,b) - K_p - K_\ell]$$

$$= \underset{p}{E}\,\underset{\ell}{E}\,\max_b \pi(p,\ell,b) - \bar{K}_p - \bar{K}_\ell$$

Therefore we should pay to learn p and ℓ simultaneously when

$$\bar{K}_p + \bar{K}_\ell < V^N_{p\ell} = (\underset{p}{E}\,\underset{\ell}{E}\,\max_b - \max_b\,\underset{p}{E}\,\underset{\ell}{E})\ \pi(p,\ell,b) = 29/96$$

Now consider the problem of buying the two pieces of information sequentially. Suppose we have already paid K_p to learn p. At the same time that we learned p, we also found out how much the information cost. Next we must decide whether or not to pay K_ℓ for ℓ. If we decide not to pay for ℓ, our expected profit as a function of p and K_p is

$$\max_b \underset{\ell}{E}\,\pi(p,\ell,b) - K_p$$

If we decide to pay for ℓ, our expected profit as a function of p and K_p is

$$\underset{\ell}{E}\,\underset{K_\ell}{E}\,[\max_b \pi(p,\ell,b) - K_p - K_\ell] = \underset{\ell}{E}\,\max_b \pi(p,\ell,b) - K_p - K_\ell$$

We should pay K_ℓ for ℓ when

$$\bar{K}_\ell < (\underset{\ell}{E}\ \max_b - \max_b\ \underset{\ell}{E})\ \pi(p,\ell,b) = (1 - p/2)^2/2$$

This is the same decision rule that we found for the case of certain prices, except now K_ℓ is replaced by $\bar{K}_\ell$. Carrying out the same calculations that we did in Chapter 2 we find that our expected profit after learning p and K_p is

$$\max\left\{\begin{matrix} \underset{\ell}{E}\ \max_b \pi(p,\ell,b) - \bar{K}_\ell \\ \\ \max_b \underset{\ell}{E}\ \pi(p,\ell,b) \end{matrix}\right\} - K_p .$$

Before we know p and K_p our expected profit is

$$\underset{p}{E}\ \underset{K_p}{E}\left[\max\left\{\begin{matrix} \underset{\ell}{E}\ \max_b \pi(p,\ell,b) - \bar{K}_\ell \\ \\ \max_b \underset{\ell}{E}\ \pi(p,\ell,b) \end{matrix}\right\} - K_p\right]$$

$$= \underset{p}{E}\ \max\left\{\begin{matrix} \underset{\ell}{E}\ \max_b \pi(p,\ell,b) - \bar{K}_\ell \\ \\ \max_b \underset{\ell}{E}\ \pi(p,\ell,b) \end{matrix}\right\} - \bar{K}_p$$

Thus we should pay K_p to learn p, with an option to pay K_ℓ for ℓ, when

$$\bar{K}_p < V_p(\bar{K}_\ell) = E\ \max\left\{\begin{matrix} \underset{\ell}{E}\ \max_b \pi(p,\ell,b) - \bar{K}_\ell \\ \\ \max_b \underset{\ell}{E}\ \pi(p,\ell,b) \end{matrix}\right\} - \max_b \underset{p}{E}\ \underset{\ell}{E}\ \pi(p,\ell,b)$$

$$= \left\{\begin{matrix} 29/96 - \bar{K}_\ell : \bar{K}_\ell \le 1/8 \\ 33/96 + (/43)\bar{K}_\ell\sqrt{2\bar{K}_\ell} - 2\bar{K}_\ell : 1/8 < \bar{K}_\ell < 1/2 \\ 1/96 : 1/2 \le \bar{K}_\ell \end{matrix}\right\}$$

We can use the same reasoning to show that we should pay K_ℓ to learn ℓ, with an option to pay K_p for p, whenever

$$\bar{K}_\ell < V_\ell(K_p) = \underset{\ell}{E} \max \begin{Bmatrix} \underset{p}{E} \max_b \pi(p,\ell,b) - \bar{K}_p \\ \\ \max_b \underset{p}{E} \pi(p,\ell,b) \end{Bmatrix} - \max_b \underset{p}{E} \underset{\ell}{E} \pi(p,\ell,b)$$

$$= \begin{Bmatrix} 29/96 + (2/3)\bar{K}_p\sqrt{2\bar{K}_p} - \bar{K}_p/2 : \bar{K}_p < 1/8 \\ 27/96 : \bar{K}_p \le 1/8 \end{Bmatrix}$$

These decision rules for buying perfect information about p and ℓ are the same as the ones we found in Chapter 2 for certain prices, except this time K_p and K_ℓ are replaced by their expected values. Therefore, regardless of the probability density functions for K_p and K_ℓ, we can base our decisions on $\bar{K}_p$ and $\bar{K}_\ell$ as long as all the random variables are independent. In fact we can draw a price diagram similar to the one in Chapter 2 with $\bar{K}_p$ and $\bar{K}_\ell$ measured along the axes. This is shown in Fig. 5.1. The decision regions in Fig. 5.1 show us what to do for any pair of expected prices, $(\bar{K}_p,\bar{K}_\ell)$.

Now we will relax the assumption that K_p and K_ℓ are independent. Assume that p and ℓ are still independent of each other and the prices, but that K_p and K_ℓ are dependent. Thus

$$\{p,\ell,K_p,K_\ell|\mathcal{S}\} = \{p|\mathcal{S}\}\ \{\ell|\mathcal{S}\}\ \{K_p,K_\ell|\mathcal{S}\}$$

The probability density functions for p and ℓ are the same as before, and $\{K_p,K_\ell|\mathcal{S}\}$ is arbitrary.

Using the same logic that we did in the case where K_p and K_ℓ

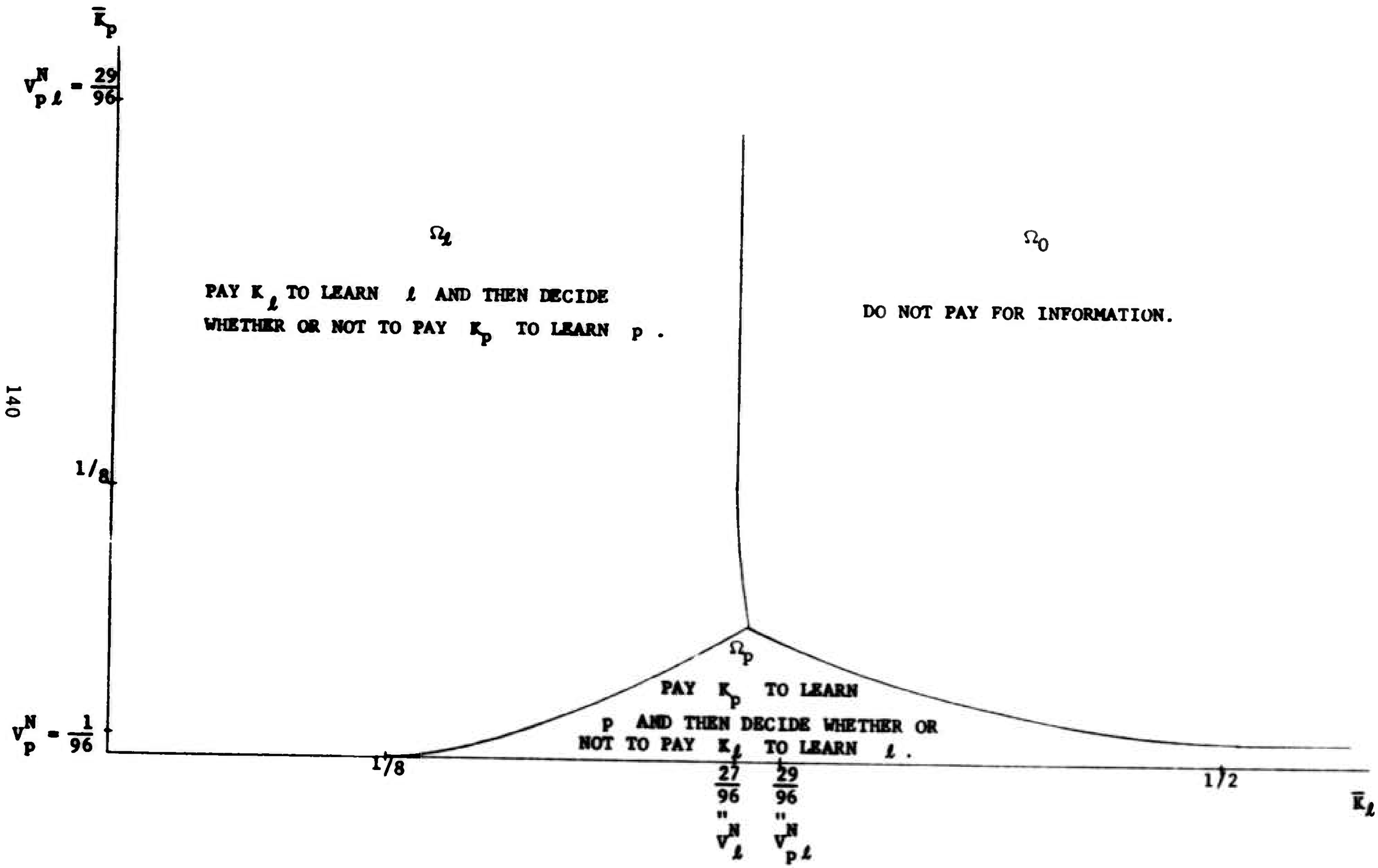


Figure 5.1. Decision regions for uncertain, independent prices

were independent, we find that we should pay K_p to learn p by itself when

$$\bar{K}_p < V_p^N = (\underset{p}{E} \max_b \underset{\ell}{E} - \max_b \underset{p}{E} \underset{\ell}{E}) \; \pi(p,\ell,b) = 1/96$$

Similarly we should pay K_ℓ to learn ℓ by itself when

$$\bar{K}_\ell < V_\ell^N = (\underset{\ell}{E} \max_b \underset{p}{E} - \max_b \underset{p}{E} \underset{\ell}{E}) \; \pi(p,\ell,b) = 27/96$$

We should pay K_p plus K_ℓ to learn both state variables when

$$\bar{K}_p + \bar{K}_\ell < V_{p\ell}^N = (\underset{p}{E} \underset{\ell}{E} \max_b - \max_b \underset{p}{E} \underset{\ell}{E}) \; \pi(p,\ell,b) = 29/96$$

So far the fact that K_p and K_ℓ are dependent has not changed the decision rules. However when we consider sequential information the decision rules are different. Suppose we have already paid K_p to learn p, and we are trying to decide whether or not to pay K_ℓ for ℓ. If we decide not to pay for ℓ, our expected profit as a function of p and K_p is

$$\max_b \underset{\ell}{E} \; \pi(p,\ell,b) - K_p$$

If we decide to pay for ℓ, our expected profit is

$$\underset{K_p}{D} \; \underset{\ell}{E} \; \underset{K_\ell}{E} [\max \pi(p,\ell,b) - K_p - K_\ell] = \underset{\ell}{E} \max_b \pi(p,\ell,b) - K_p - \underset{K_p}{D} \; \underset{K_\ell}{E} K_\ell$$

We should pay K_ℓ for ℓ whenever

$$\underset{K_p}{D} \; \underset{K_\ell}{E} K_\ell < (\underset{\ell}{E} \max_b - \max_b \underset{\ell}{E}) \; \pi(p,\ell,b)$$

This is not the same decision rule that we found previously. The left side of this inequality is a function of K_p, and is not equal to $\bar{K}_\ell$.

The quantity $\underset{K_p}{D}\ \underset{K_\ell}{E}\ K_\ell$ is determined by the dependency of K_ℓ on K_p, or, in other words, it is determined by $\{K_p, K_\ell | \mathcal{E}\}$. Since there is no way to describe allof the possible dependencies between K_p and K_ℓ -- or all of the possible joint probability density functions for K_p and K_ℓ -- by a single number, it is impossible to represent $\underset{K_p}{D}\ \underset{K_\ell}{E}\ K_\ell$ as a single point in the $\bar{K}_p$-$\bar{K}_\ell$ diagram.

Continuing the calculations as before, we find that our expected profit after learning p and K_p is

$$\max\left\{\begin{array}{l} \underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) - \underset{K_p}{D}\ \underset{K_\ell}{E}\ K_\ell \\ \\ \underset{b}{\max}\ \underset{\ell}{E}\ \pi(p,\ell,b) \end{array}\right\} - K_p$$

$$= \underset{K_p}{D}\ \max\left\{\begin{array}{l} \underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) - \underset{K_\ell}{E}\ K_\ell \\ \\ \underset{b}{\max}\ \underset{\ell}{E}\ \pi(p,\ell,b) \end{array}\right\} - K_p$$

Before we know p and K_p our expected profit is

$$\underset{p}{E}\ \underset{K_p}{E}\ \max\left\{\begin{array}{l} \underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) - \underset{K_\ell}{E}\ K_\ell \\ \\ \underset{b}{\max}\ \underset{\ell}{E}\ \pi(p,\ell,b) \end{array}\right\} - \bar{K}_p$$

Thus we should pay K_p to learn p, with an option to pay K_ℓ for ℓ, whenever

$$\bar{K}_p < \underset{p}{E}\ \underset{K_p}{E}\ \max\left\{\begin{array}{l} \underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b) - \underset{K_\ell}{E}\ K_\ell \\ \\ \underset{b}{\max}\ \underset{\ell}{E}\ \pi(p,\ell,b) \end{array}\right\} - \underset{p}{E}\ \underset{\ell}{E}\ \underset{b}{\max}\ \pi(p,\ell,b)$$

$$= \left\{\begin{array}{l} 29/96 - \bar{K}_\ell : \underset{K_p}{D}\ \underset{K_\ell}{E}\ K_\ell \le 1/8 \\ \underset{K_p}{E}\ [33/96 + (4/3)(\underset{K_\ell}{E}\ K_\ell)\sqrt{2(\underset{K_\ell}{E}\ K_\ell)} - 2(\underset{K_\ell}{E}\ K_\ell)] : 1/8 < \underset{K_p}{D}\ \underset{K_\ell}{E}\ K_\ell < 1/2 \\ 1/96 : 1/2 \le \underset{K_p}{D}\ \underset{K_\ell}{E}\ K_\ell \end{array}\right\}$$

Similarly we can show that we should pay K_ℓ to learn ℓ, with an option to pay K_p for p, whenever

$$\bar{K}_\ell < \underset{\ell}{E}\ \underset{K_\ell}{E}\ \max\left\{\begin{array}{l}\underset{p}{E}\ \max_b \pi(p,\ell,b) - \underset{K_p}{E}\ K_p \\ \\ \max_b \underset{\ell}{E}\ \pi(p,\ell,b)\end{array}\right\} - \underset{p}{E}\ \underset{\ell}{E}\ \max_b \pi(p,\ell,b)$$

$$= \left\{\begin{array}{ll}\underset{K_\ell}{E}\ [29/96 + (4/3)(\underset{K_p}{E}\ K_p)\sqrt{2(\underset{K_p}{E}\ K_p)} - 2(\underset{K_p}{E}\ K_p)] & : \underset{K_\ell}{D}\ \underset{K_p}{E}\ K_p < 1/8 \\ \\ 27/96 & : \underset{K_\ell}{D}\ \underset{K_p}{E}\ K_p \geq 1/8\end{array}\right\}$$

Obviously the analysis developed in the previous chapters breaks down when K_p and K_ℓ are dependent. It is no longer possible to define decision regions in the $\bar{K}_p - \bar{K}_\ell$ space, or define the value of sequential information as a function of the expected prices. We could define the value of sequential information in terms of the joint probability density function for K_p and K_ℓ

$$V_p = V_p(\{K_p, K_\ell | \mathcal{E}\})$$

However this sort of function is not very useful. Since the dependencies between K_p and K_ℓ, as given by the joint probability density function, cannot be represented by a finite set of numbers, it is impossible to describe decision regions in any finite-dimensional Euclidean space. However, for any given joint probability density function, $\{K_p, K_\ell | \mathcal{E}\}$, we can solve the equations above to find the best way to purchase information sequentially.

General Formulation of Sequential Information Problems with Uncertain Prices

In a general decision problem with sequential information and uncertain prices, we have a profit π that depends on a control variable c and a set of state variables $(x_1,\ldots,x_m)$. We have an opportunity to learn any of a set of observables $(y_1,\ldots,y_n)$ for prices $(K_{y_1},\ldots,K_{y_n})$, respectively. The state variables, observables, and prices are all random variables, and our state of information about these random variables is given by our assessment of the joint probability density function

$$\{x_1,\ldots,x_m,y_1,\ldots,y_n,K_{y_1},\ldots,K_{y_n} \mid \mathcal{E}\}$$

When we commit ourselves to paying for the i^{th} observable, we learn the actual value of both y_i and K_{y_i} . Then we pay K_{y_i} for the information. The complete decision tree for an expected-value decision maker is shown in Fig. 5.2.

The expected value of learning y_i sequentially, when we are trying to maximize expected profit, and when the cost of information is uncertain, is

$$\underset{y_i}{E}\ \underset{K_{y_i}}{E}\ \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi(x_1,\ldots,x_m,c) \\ \max\limits_{j\neq i} \left(\underset{y_j}{E}\ \underset{K_{y_j}}{E} \left[\max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E}\ \pi(x_1,\ldots,x_m,c) \\ \max\limits_{k\neq i,j} \left(\underset{y_j}{E}\ \underset{K_{y_j}}{E}\ [\ldots] \right) \end{array} \right\} - K_{y_j} \right] \right) \end{array} \right\}$$

$$- \max_{c} \underset{X}{E} \cdots \underset{x}{E}\ \pi(x_1,\ldots,x_m,c)$$

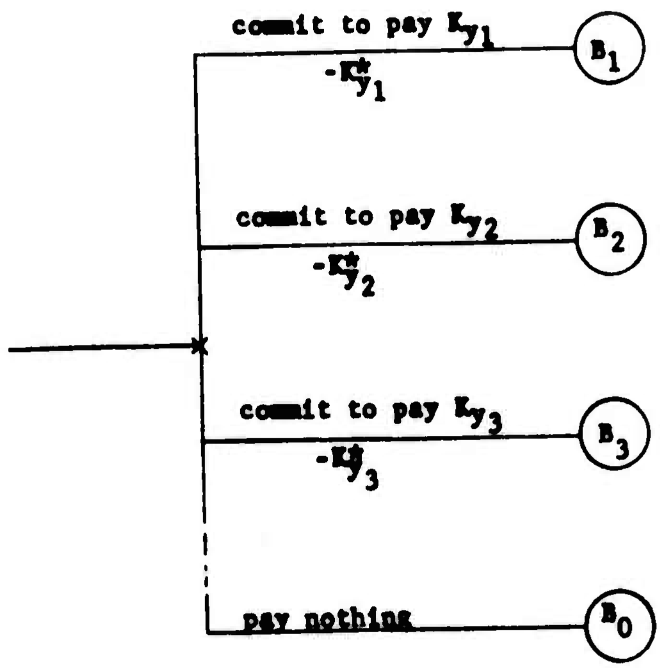


Figure 5.2a. Sequential information decision tree with uncertain prices

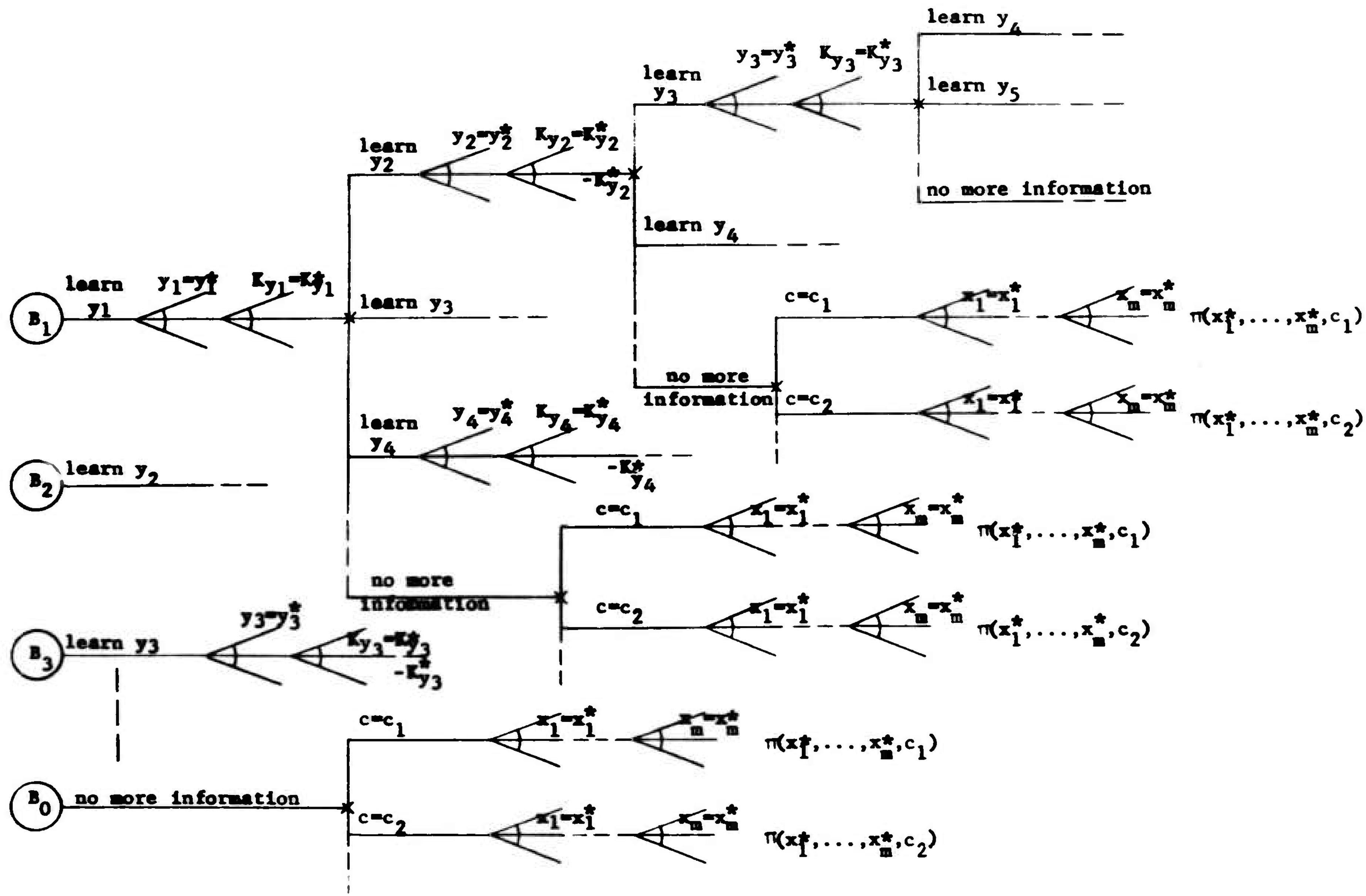


Figure 5.2b. Sequential information decision tree with uncertain prices (continued)

$$= \underset{y_i}{E}\,\underset{K_{y_i}}{E}\begin{cases} \max\limits_c \underset{x_1}{E}\cdots\underset{x_m}{E}\,\pi(x_1,\ldots,x_m,c) : \underset{y_i}{D}\,\underset{K_{y_i}}{D}\,\max\limits_c \underset{x_1}{E}\cdots\underset{x_m}{E} \\ \qquad\qquad \pi(x_1,\ldots,x_m,c) \geq \underset{y_i}{D}\,\underset{K_{y_i}}{D}\,\max\limits_{j\neq i}(\underset{y_j}{E}\,\underset{K_{y_j}}{E}\,[\ldots]) \\ \max\limits_{j\neq i}(\underset{y_j}{E}\,\underset{K_{y_j}}{E}\,[\max\{\ldots\} - K_{y_j}]) : \text{otherwise} \end{cases}$$

The question of whether or not the following inequality holds

$$\underset{y_i}{D}\,\underset{K_{y_i}}{D}\,\max_c \underset{x_1}{E}\cdots\underset{x_m}{E}\,\pi(x_1,\ldots,x_m,c)$$

$$\geq \underset{y_i}{D}\,\underset{K_{y_i}}{D}\,\max_{j\neq i}(\underset{y_j}{E}\,\underset{K_{y_j}}{E}\,[\max\{\ldots\} - K_{y_j}])$$

depends on $\underset{y_i}{D}\,\underset{K_{y_i}}{D}\,\underset{K_{y_j}}{E}\,K_{y_j}$ rather than on $\bar{K}_{y_j}$. If K_{y_j} is independent of y_i and K_{y_i}, then $\underset{y_i}{D}\,\underset{K_{y_i}}{D}\,\underset{K_{y_j}}{E}\,K_{y_j}$ is the same as $\bar{K}_{y_j}$ and the decision rule depends on $\bar{K}_{y_j}$. However, if K_{y_j} is not independent of y_i and K_{y_i}, the decision rule will depend on $\underset{y_i}{D}\,\underset{K_y}{D}\,\underset{K_y}{E}\,K_{y_j}$, which requires a knowledge of the entire joint probability density function rather than just the expected values of the prices. Therefore we can determine the value of sequential information and the decision regions in terms of the expected values of the prices whenever K_{y_j} is independent of y_i and K_{y_i}, for all $j = 1,\ldots,n$ and for all i not equal to j. When the prices are not independent--or when an observable is not independent of the price of another observable--the analysis breaks down However we can still solve the problem using the decision tree in Fig. 5.2 if the joint probability density function for all the random variables is known.

The Bidding Problem with Non-Additive Prices

Now we can turn our attention to the case where the prices are certain but they are not additive. This means that the cost of learning two or more of the observables need not equal the sum of the costs of learning each individually. For example, in the bidding problem K_p is the cost of learning p, and K_ℓ is the cost of learning ℓ. We assumed that the total cost of learning both p and ℓ, either simultaneously or sequentially, was K_p plus K_ℓ. We are now going to relax that assumption and define the following prices:

$K_{p\ell}$ = cost of learning p and ℓ simultaneously

$K_{p|\ell}$ = cost of learning p sequentially after we already know ℓ

$K_{\ell|p}$ = cost of learning ℓ sequentially after we already know p.

Any time we change the pricing structure, we must change the decision rules for buying information because the rules are described in terms of the prices. For the bidding problem with non-additive prices, the decision rules depend on all five prices $(K_p, K_\ell, K_{p\ell}, K_{p|\ell},$ and $K_{\ell|p})$ rather than on just K_p and K_ℓ. Thus the price diagram for this problem will have five dimensions, with one of the prices measured along each axis. Although it is difficult to visualize decision regions in five dimensions, the regions are fairly easy to describe algebraically, and we can determine them with the procedures used in the previous chapters.

If we do not receive any information about p or ℓ, our expected profit is still

$$\max_b \underset{p}{E} \underset{\ell}{E} \pi(p,\ell,b) = 27/96$$

The reasoning we used before shows that we should pay K_p for p by itself when

$$K_p < V_p^N = (\underset{p}{E}\,\underset{b}{\max}\,\underset{\ell}{E} - \underset{b}{\max}\,\underset{p}{E}\,\underset{\ell}{E})\ \pi(p,\ell,b) = 1/96$$

Similarly, we should pay K_ℓ for ℓ by itself when

$$K_\ell < V_\ell^N = (\underset{\ell}{E}\,\underset{b}{\max}\,\underset{p}{E} - \underset{b}{\max}\,\underset{p}{E}\,\underset{\ell}{E})\ \pi(p,\ell,b) = 27/96$$

If we pay $K_{p\ell}$ to learn p and ℓ simultaneously, our expected profit before receiving the information is

$$\underset{p}{E}\,\underset{\ell}{E}\,\underset{b}{\max}\ \pi(p,\ell,b) - K_{p\ell}$$

Thus we should pay to learn p and ℓ simultaneously whenever

$$K_{p\ell} < V_{p\ell}^N = (\underset{p}{E}\,\underset{\ell}{E}\,\underset{b}{\max} - \underset{b}{\max}\,\underset{p}{E}\,\underset{\ell}{E})\ \pi(p,\ell,b) = 29/96$$

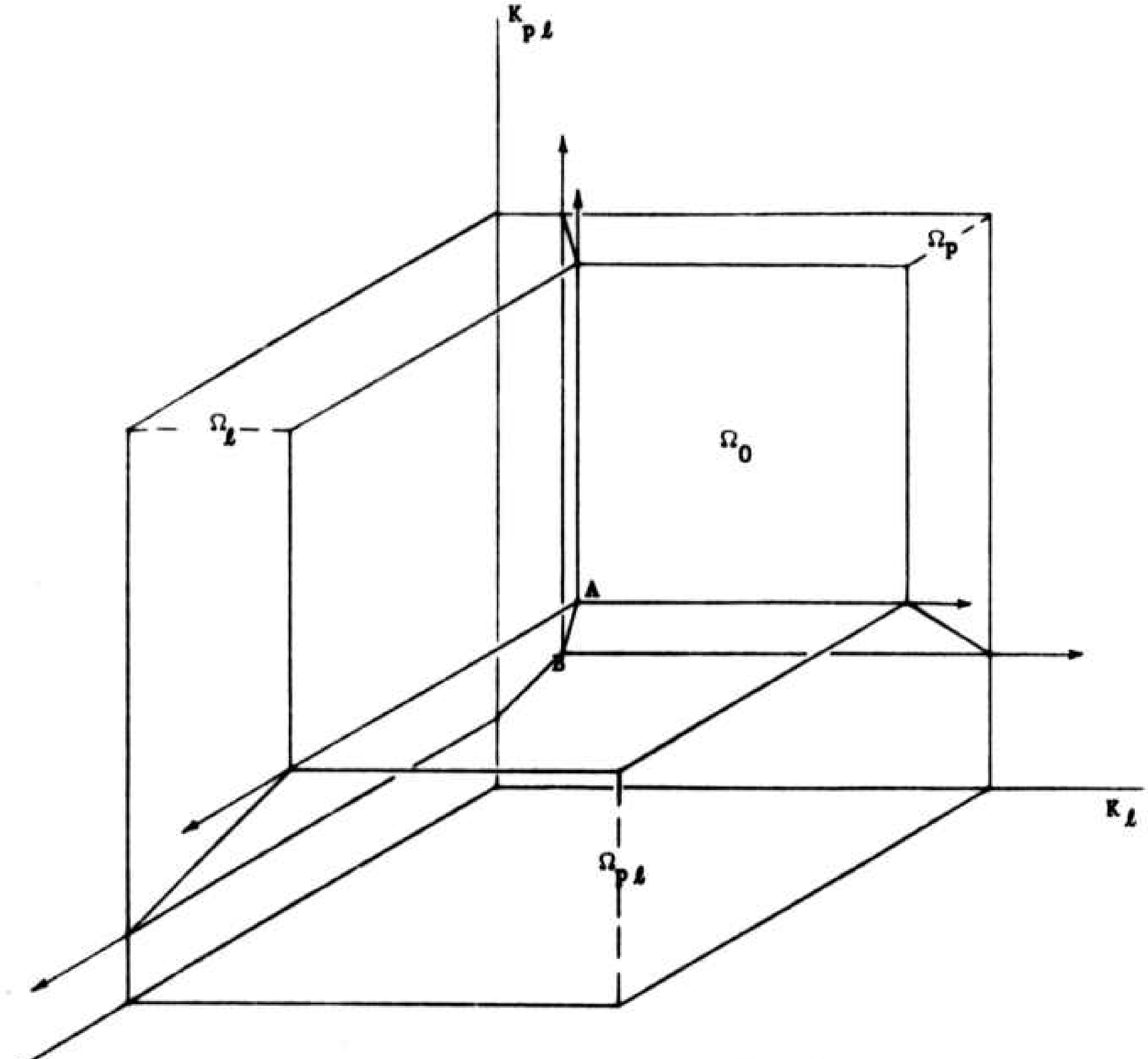
If we do not have the option of learning the state variables sequentially, the price diagram would have three dimensions corresponding to K_p, K_ℓ, and $K_{p\ell}$. Our best decision is to pay K_p for p when

$$V_p^N - K_p > 0$$

$$V_p^N - K_p > V_\ell^N - K_\ell$$

$$V_p^N - K_p > V_{p\ell}^N - K_{p\ell}$$

Similar inequalities describe the sets of prices $(K_p, K_\ell, K_{p\ell})$ where our best initial decision is to buy ℓ, or to buy both state variables simultaneously. These decision regions are shown in Fig. 5.3.



Ω_0 = buy nothing

Ω_p = buy p

Ω_ℓ = buy ℓ

$\Omega_{p\ell}$ = buy both p and ℓ

$A = (V_p^N, V_\ell^N, V_{p\ell}^N)$

$B = (0, V_\ell^N - V_p^N, V_{p\ell}^N - V_p^N)$

$C = (0, 0, V_{p\ell}^N - V_\ell^N)$

(Drawn assuming $V_{p\ell}^N > V_\ell^N > V_p^N$; not drawn to scale)

Figure 5.3. Decision regions for individual and simultaneous information when prices are not additive.

Now consider the possibility of sequential information. Suppose we have already paid K_p to learn p, and we are trying to decide whether or not to pay $K_{\ell|p}$ to learn ℓ. If we decide not to learn ℓ, our expected profit is

$$\max_b \mathop{E}_{\ell} \pi(p,\ell,b) - K_p$$

If we decide to pay for ℓ, our expected profit before receiving the information is

$$\mathop{E}_{\ell} \max_b \pi(p,\ell,b) - K_p - K_{\ell|p}$$

Since we will choose the larger of these two quantities, our expected profit prior to learning p is

$$\mathop{E}_{p} \max \begin{Bmatrix} \mathop{E}_{\ell} \max_b \pi(p,\ell,b) - K_{\ell|p} \\ \\ \max_b \mathop{E}_{\ell} \pi(p,\ell,b) \end{Bmatrix} - K_p$$

Thus we should pay K_p to learn p, with an option to pay $K_{\ell|p}$ for ℓ, whenever

$$K_p < V_p(K_{\ell|p}) = \mathop{E}_{p} \max \begin{Bmatrix} \max_b \mathop{E}_{\ell} \pi(p,\ell,b) - K_{\ell|p} \\ \\ \mathop{E}_{\ell} \max_b \pi(p,\ell,b) \end{Bmatrix} - \max_b \mathop{E}_{p} \mathop{E}_{\ell} \pi(p,\ell,b)$$

$$= \begin{Bmatrix} 29/96 - K_{\ell|p} : K_{\ell|p} \le 1/8 \\ 33/96 + (4/3)K_{\ell|p}\sqrt{2K_{\ell|p}} - 2K_{\ell|p} : 1/8 < K_{\ell|p} < 1/2 \\ 1/96 : 1/2 \le K_{\ell|p} \end{Bmatrix}$$

This is the same function that we derived in Chapter 2, except that K_ℓ

has been replaced by $K_{\ell|p}$.

The same reasoning shows that we should pay K_ℓ to learn ℓ , with an option to pay $K_{p|\ell}$ for p , whenever

$$K_\ell < V_\ell(K_{p|\ell}) = \mathop{E}_{\ell} \max\left\{\begin{matrix} \mathop{E}_{p} \max_{b} \pi(p,\ell,b) - K_{p|\ell} \\ \\ \max_{b} \mathop{E}_{p} \pi(p,\ell,b) \end{matrix}\right\} - \max_{b} \mathop{E}_{p} \mathop{E}_{\ell} \pi(p,\ell,b)$$

$$= \left\{\begin{matrix} 29/96 + (2/3)K_{p|\ell}\sqrt{2K_{p|\ell}} - K_{p|\ell}/2 : K_{p|\ell} < 1/8 \\ \\ 27/96 : K_{p|\ell} \geq 1/8 \end{matrix}\right\}$$

In order to maximize our expected profit we should pay K_p to learn p , with an option to pay $K_{\ell|p}$ for ℓ , whenever

$$V_p(K_{\ell|p}) - K_p > 0$$

$$V_p(K_{\ell|p}) - K_p > V_\ell^N - K_\ell$$

$$V_p(K_{\ell|p}) - K_p > V_{p\ell}^N - K_{p\ell}$$

$$V_p(K_{\ell|p}) - K_p > V_\ell(K_{p|\ell}) - V_\ell$$

(It is not necessary to add the condition

$$V_p(K_{\ell|p}) - K_p > V_p^N - K_p$$

since we can show that $V_p(K_{\ell|p}) > V_p^N$ for any value of $K_{\ell|p}$.) Similar inequalities describe the set of prices $(K_p, K_\ell, K_{p\ell}, K_{p|\ell}, K_{\ell|p})$ such that our best initial decision is to buy ℓ with an option to buy p , to buy either state variable individually, or to buy both

simultaneously. We will not attempt to visualize the five-dimensional price diagram.

Unless we have some method of restricting the five prices so that they can be described with less than five numbers, we are forced to consider decision regions in spaces with five dimensions. One way to restrict the prices is to assume that they are additive. Another way, which allows us to reduce the cost of one observable if we have already learned another, is to assume that there are reduction factors λ_p and λ_ℓ such that the cost of learning ℓ is $\lambda_\ell K_\ell$ when we have already paid K_p to learn p. Similarly, the cost of learning p is $\lambda_p K_p$ when we have already paid K_ℓ for ℓ. Normally λ_p and λ_ℓ will both lie between zero and one, but we allow them to have any positive values. When λ_p and λ_ℓ both equal one, we have the case of additive prices.

Consider the bidding problem for arbitrary λ_p and λ_ℓ. The reasoning we used previously shows that we should pay K_p to learn p by itself when

$$K_p < V_p^N = (\underset{p}{E} \max_b \underset{\ell}{E} - \max_b \underset{p}{E} \underset{\ell}{E}) \pi(p,\ell,b) = 1/96$$

Similarly we should pay K_ℓ to learn ℓ by itself when

$$K_\ell < V_\ell^N = (\underset{\ell}{E} \max_b \underset{p}{E} - \max_b \underset{p}{E} \underset{\ell}{E}) \pi(p,\ell,b) = 27/96$$

If we decide to learn p and ℓ simultaneously, we have a choice of paying $(K_p + \lambda_\ell K_\ell)$ or $(K_\ell + \lambda_p K_p)$. The first price corresponds to first learning p and then ℓ. The second price corresponds to

first learning ℓ and then p . We get the same information either way since we will receive both pieces of information simultaneously. Assuming that we pay the lower of the two prices, our expected profit after deciding to pay for p and ℓ , but before we receive the information, is

$$\underset{p}{E}\,\underset{\ell}{E}\,\max\ \pi(p,\ell,b) - \min\begin{Bmatrix} K_p + \lambda_\ell K_\ell \\ \\ K_\ell + \lambda_p K_p \end{Bmatrix}$$

Thus we should pay to learn p and ℓ simultaneously whenever

$$\min\begin{Bmatrix} K_p + \lambda_\ell K_\ell \\ \\ K_\ell + \lambda_p K_p \end{Bmatrix} < V^N = (\underset{p}{E}\,\underset{\ell}{E}\,\underset{b}{\max} - \underset{b}{\max}\,\underset{p}{E}\,\underset{\ell}{E})\ \pi(p,\ell,b)$$

The decision rules for individual and simultaneous information about p and ℓ are shown in Fig. 5.4. The regions where we will buy perfect information about p or ℓ individually are the sets of price pairs represented by points below the line A-B and to the left of the line C-D, respectively. These are the same as the sets of price pairs that we found in Chapter 2 for additive prices. (See Fig. 2.2.) However we are now willing to learn p and ℓ simultaneously for any pair of prices represented by a point below or the the left of the boundary G-I-H. When the prices are additive, a single 45° line bounds the pairs of prices such that we will pay to learn p and ℓ simultaneously. In Fig. 5.4 the 45° line splits into two lines, E-H and G-F. We are willing to pay any pair of prices, with the appropriate reductions, represented by a point below or to the left of either line.

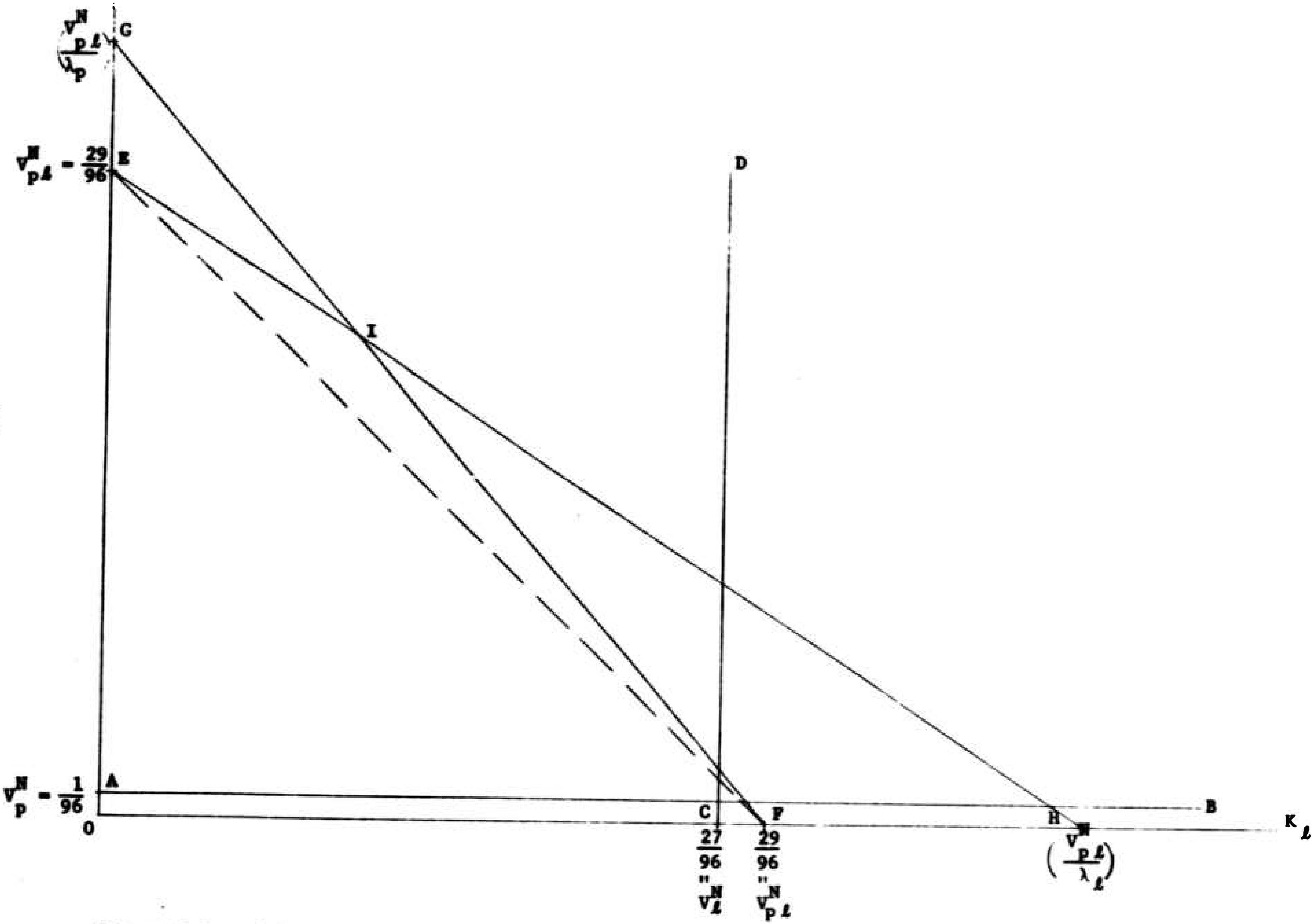


Figure 5.4. Price diagram for individual and simultaneous information with reduced prices

Now consider the case of sequential information when the prices are altered by the reduction factors, λ_p and λ_ℓ. Suppose we have already paid K_p to learn p, and we are trying to decide whether or not to pay $\lambda_\ell K_\ell$ to learn ℓ. If we decide not to learn ℓ, our expected profit is

$$\max_b \underset{\ell}{E}\, \pi(p,\ell,b) - K_p$$

If we decide to pay for ℓ, our expected profit prior to receiving the information is

$$\underset{\ell}{E} \max_b \pi(p,\ell,b) - K_p - \lambda_\ell K_\ell$$

Since we will choose the larger of these two quantities, our expected profit prior to learning p is

$$\underset{p}{E} \max \begin{Bmatrix} \underset{\ell}{E} \max\limits_b \pi(p,\ell,b) - \lambda_\ell K_\ell \\ \\ \max\limits_b \underset{\ell}{E}\, \pi(p,\ell,b) \end{Bmatrix} - K_p$$

Thus we should pay K_p to learn p, with an option to pay $\lambda_\ell K_\ell$ for ℓ, when

$$K_p < V_p(\lambda_\ell K_\ell) = \underset{p}{E} \max \begin{Bmatrix} \max\limits_b \underset{\ell}{E}\, \pi(p,\ell,b) - \lambda_\ell K_\ell \\ \\ \underset{\ell}{E} \max\limits_b \pi(p,\ell,b) \end{Bmatrix} - \max_b \underset{p}{E}\, \underset{\ell}{E}\, \pi(p,\ell,b)$$

$$= \begin{Bmatrix} 29/96 - \lambda_\ell K_\ell : K_\ell \le 1/(8\lambda_\ell) \\ 33/96 + (4/3)\lambda_\ell K_\ell \sqrt{2\lambda_\ell K_\ell} - 2\lambda_\ell K_\ell : 1/(8\lambda_\ell) < K_\ell < 1/(2\lambda_\ell) \\ 1/96 : 1/(2\lambda_\ell) \le K_\ell \end{Bmatrix}$$

Similarly, we should pay K_ℓ to learn ℓ, with an option to pay $\lambda_p K_p$ for p, when

$$K_\ell < V_\ell(\lambda_p K_p) = \underset{\ell}{E} \max \begin{Bmatrix} \underset{p}{E} \max\limits_b \pi(p,\ell,b) - \lambda_p K_p \\ \\ \max\limits_b \underset{p}{E} \pi(p,\ell,b) \end{Bmatrix} - \max_b \underset{p}{E}\, \underset{\ell}{E}\, \pi(p,\ell,b)$$

$$= \begin{Bmatrix} 29/96 + (2/3)\lambda_p K_p \sqrt{2\lambda_p K_p} - \lambda_p K_p/2 : K_p < 1/(8\lambda_p) \\ \\ 27/96 : K_p \geq 1/(8\lambda_p) \end{Bmatrix}$$

These decision rules are shown graphically in Fig. 5.5. The effect of λ_ℓ is to stretch the V_p boundary for additive prices (dashed curve A-B'-H'-C'-D in Fig. 5.5) to the right by a factor of $(1/\lambda_\ell)$. Similarly, the effect of λ_p is to stretch the V_ℓ boundary for additive prices (dashed curve E-H'-F'-G in Fig. 5.5) up by a factor of $(1/\lambda_p)$. Thus it is possible to visualize the decision regions in a two-dimensional space, where λ_p and λ_ℓ are used to stretch the decision boundaries for the corresponding problem with additive prices.

General Formulation of Sequential-Information Problems with Non-Additive Prices

As before, the general problem with sequential information is characterized by a profit function π that depends on a control variable c and a set of state variables $(x_1,\ldots,x_m)$. This time the observables $(y_1,\ldots,y_n)$ have certain prices, but the prices depend on the order in which the observables are purchased. We can define all of the possible prices as follows:

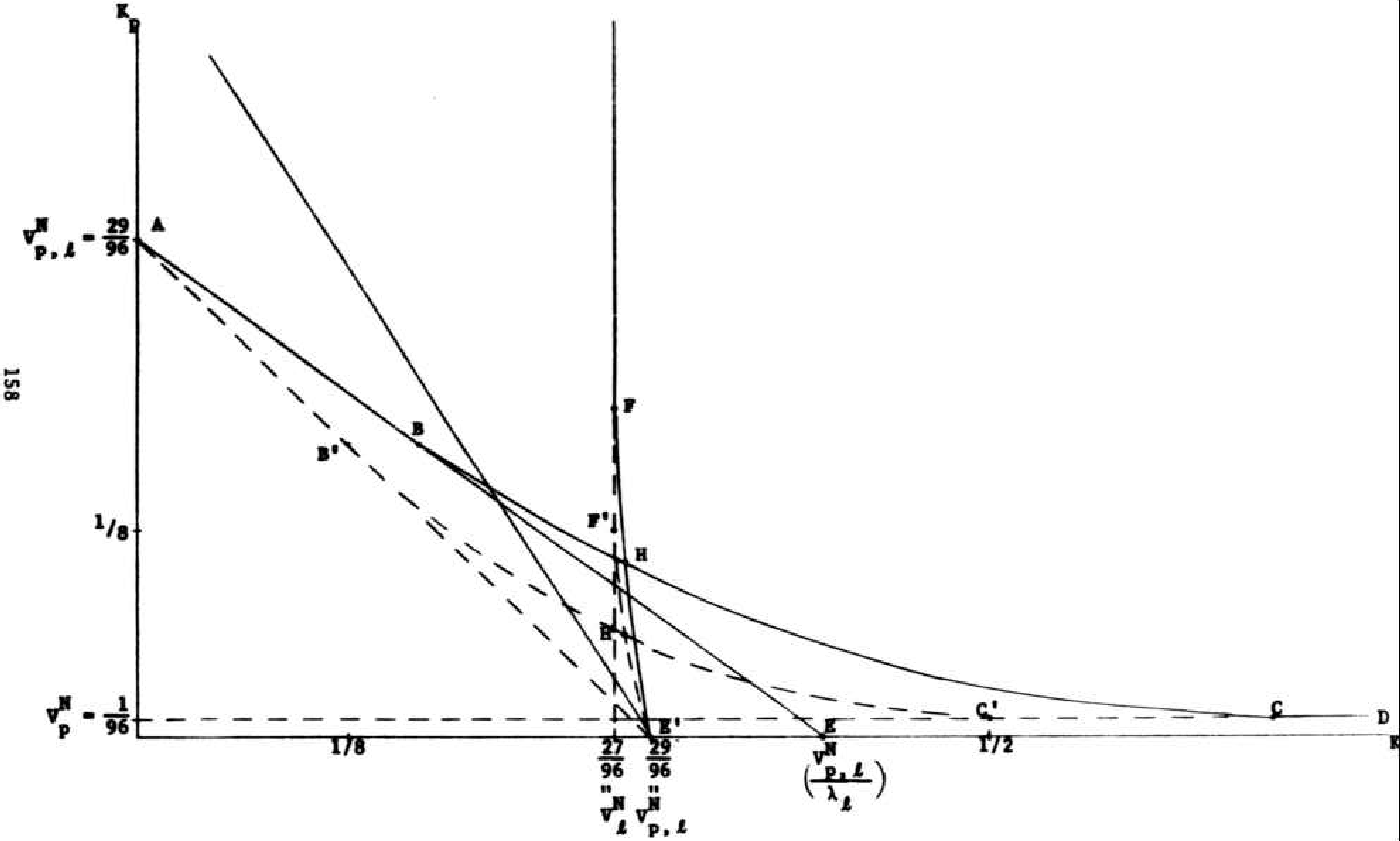


Figure 5.5. Price diagram for sequential information with reduced prices

K_{y_i} = cost of learning y_i when none of the other observables are known;

$K_{y_i|y_j}$ = cost of learning y_i when y_j is known;

$K_{y_i|y_j y_k \cdots y_r}$ = cost of learning y_i when y_j , y_k , etc. are known;

$K_{y_i y_j \cdots y_r}$ = cost of learning y_i , y_j , etc. simultaneously when none of the observables are known and no additional information will be purchased.

(If we allow the decision maker to buy sets of observables simultaneously and then make sequential decisions about buying more information, we need to define even more prices to cover all of the possible sequential purchases of simultaneous blocks of information. The situation is already complicated enough, so we will only allow the decision maker to make a single simultaneous purchase with no prior or subsequent purchases of information. We would also have more prices if the cost of an observable depended on the actual value of the observables, as it might if we did destructive testing. However we will not allow this type of price either.)

It is easy to show that we will have $n2^{(n-1)}$ sequential prices (including $K_{y_1},\ldots,K_{y_n}$), and (2^n-n-1) simultaneous prices, when there are n observables. Since each of these prices represents one dimension of the price diagram, it is clear that we cannot hope to visualize the decision regions even when there are only two or three observables. However, having a large number of prices does not mean that we cannot describe the decision regions algebraically. We can determine a set of inequalities that describe each decision region by following the procedures used in the preceding chapters.

The expected value of learning y_i sequentially when we are trying to maximize expected profit and the prices of the observables are not additive, is

$$\underset{y_i}{E}\max\left\{\begin{array}{l}\max\limits_{c}\underset{x_1}{E}\cdots\underset{x_m}{E}\pi(x_1,\ldots,x_m,c)\\[2ex] \max\limits_{j\neq i}\left(\underset{y_j}{E}\max\left\{\begin{array}{l}\max\limits_{c}\underset{x_1}{E}\cdots\underset{x_m}{E}\pi(x_1,\ldots,x_m,c)\\ \max\limits_{k\neq i,j}\ (\underset{y_k}{E}\max\{\ldots\}-K_{y_k|y_iy_j})\end{array}\right\}-K_{y_j|y_i}\right)\end{array}\right\}$$

$$-\max_{c}\underset{x_1}{E}\cdots\underset{x_m}{E}\pi(x_1,\ldots,x_m,c)$$

(This is not equal to $V_{y_i}(K_{y_1|y_i},\ldots,K_{y_n|y_i})$ or $V_{y_i}(\ldots,K_{y_j|y_i},\ldots,$ $K_{y_k|y_iy_j},\ldots)$.) When we consider the subspace of the price diagram spanned by K_{y_i} , $K_{y_j|y_i}$, $K_{y_k|y_iy_j}$, etc., for some specific values of i , j , k , ..., we are able to draw decision boundaries similar to those that we found for additive prices. However the decision boundaries become much more complicated when we consider all of the dimensions of the price diagram.

The decision regions can be simplified if we restrict the prices of the observables such that they can be described with relatively few numbers. One such restriction is the case of additive prices. Another way to restrict the prices is to assume that the i^{th} observable costs K_{y_i} if it is the first piece of information that we learn, and $\lambda_{y_i}K_{y_i}$ otherwise. The reduction factor λ_{y_i} represents the savings that result from having previously set up an information-gathering process to learn a different observable. In this case we can show that the value of learning y_i sequentially, when we are trying to maximize expected profit, is

$$V_{y_i}(\lambda_{y_1}K_{y_1},\ldots,\lambda_{y_{i-1}}K_{y_{i-1}},\lambda_{y_{i+1}}K_{y_{i+1}},\ldots,\lambda_{y_n}K_{y_n})$$

$$= \underset{y_i}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m,c) \\ \max\limits_{j\neq i} \left(\underset{y_j}{E} \max \left\{ \begin{array}{l} \max\limits_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m,c) \\ \max\limits_{k\neq i,j} (\underset{y_k}{E} \max\{\ldots\} - \lambda_{y_k}K_{y_k}) \end{array} \right\} - \lambda_{y_j}K_{y_j} \right) \end{array} \right\}$$

$$- \max_{c} \underset{x_1}{E} \cdots \underset{x_m}{E} \pi(x_1,\ldots,x_m,c)$$

The corresponding decision boundaries in the price diagram are similar to those for additive prices, except they are stretched along the appropriate axes by amounts equal to the inverses of the reduction factors.

There are obviously many other ways to set up the pricing structure. Since the decision rules for buying information depend on the prices of the observables, the decision rules will be different for each price structure. In some cases it is possible to visualize the decision rules as regions in a price diagram. In other cases the large number of prices will preclude any graphical representation of the decision regions. However, as the previous examples have shown, the method of analysis is the same regardless of the pricing structure.

Summary

Under certain conditions it is possible to extend the results of the previous chapters to include the cases of uncertain and non-additive prices. When the prices of the observables are uncertain, and when they are independent of each other and of the observables, we can simply replace the price of each observable by its expected value in the decision

rules for certain prices. When the prices are not independent, it is not generally possible to describe the decision rules in terms of a finite set of numbers. However we can still use the appropriate decision tree to determine the best information-purchasing decision if we have a given joint probability density function for the state variables, the observables, and their prices.

When the prices of the observables are not additive, it is necessary to define a large number of prices to include all of the possible ways that we might purchase information. Since the dimension of the price diagram equals the number of prices, it becomes difficult to visualize the decision rules for purchasing information. However these rules are still determined by following the procedures used in the previous chapters.

Even in the cases of uncertain and non-additive prices, it is still true that the expected value of sequential information is a function of the prices of the observables or our state of information about the prices.

APPENDIX A

THE PROBABILITY DENSITY FUNCTION

OF THE PROFIT WITH SEQUENTIAL INFORMATION

In Chapter 2 we assumed that the profit and the value of perfect information were certain quantities. However the profit and the value of information are random variables since they are functions of p and ℓ. Let π_p^N, π_p^R, and π_p be random variables equal to the profit when we learn the actual value of p individually, simultaneously, and sequentially. We define π^N to be the profit when no information is received, and $\pi_{p,\ell}^N$ to be the profit when p and ℓ are learned simultaneously. (None of these random variables should be confused with $\pi(p,\ell,b)$, the profit function. As before, the superscript N indicates that no information will be purchased other than that shown in the subscript.) Let v_p^N, v_p^R, and v_p be random variables equal to the value of learning individual, simultaneous, and sequential information about p. R stands for the residual value of information when p and ℓ are learned simultaneously. The expected values of these random variables are V_p^N, V_p^R, and V_p, respectively. We define $v_{p,\ell}^N$ to be the value of learning p and ℓ simultaneously. The expected value of $v_{p\ell}^N$ is $V_{p\ell}^N$.

We can determine the probability density function for each of these random variables. This will give us an idea of the risks involved in paying to learn uncertain information. R. A. Howard has derived the probability density functions for π^N, π_p^N, v_p^N, π_ℓ^N, v_ℓ^N, $\pi_{p,\ell}^N$ and $v_{p\ell}^N$, [3]. These density functions or "lotteries" are shown in Fig. A.1. As we would expect, none of these distributions depends on the prices of

π^N = profit with no information

π_p^N = profit when we are given perfect information about p only

v_p^N = value of perfect information about p only

$\pi_{p\ell}^N$ = profit when we are given perfect information about p and ℓ simultaneously

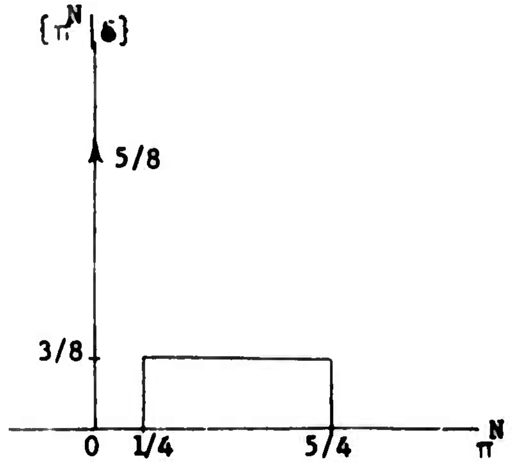


Figure A.1a

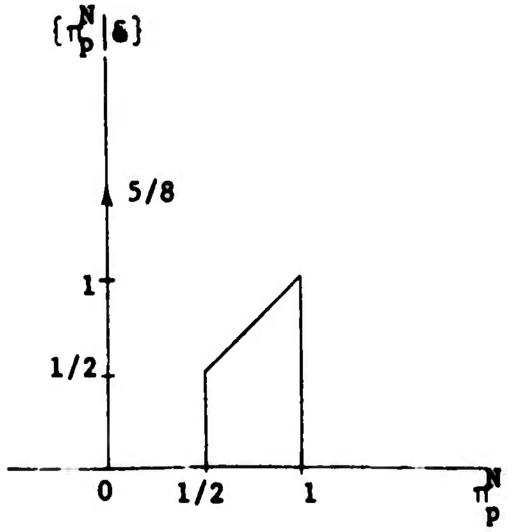


Figure A.1b

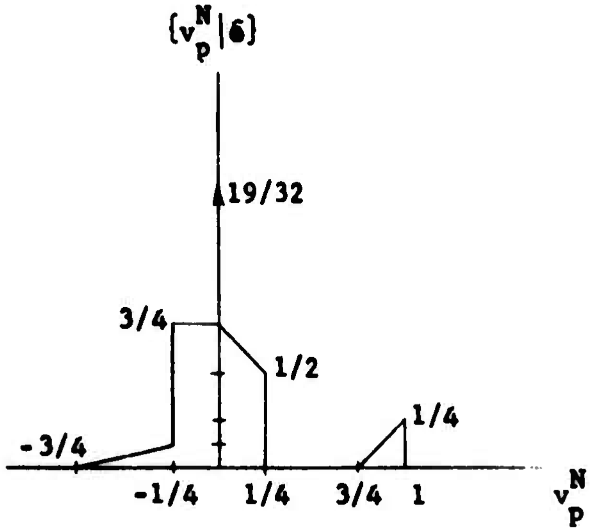


Figure A.1c

Figure A.1(a-g). Probability density functions for profit and value of information

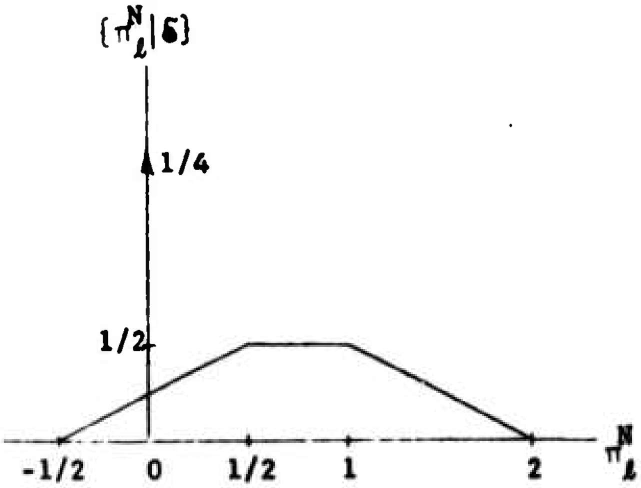


Figure A.1d.

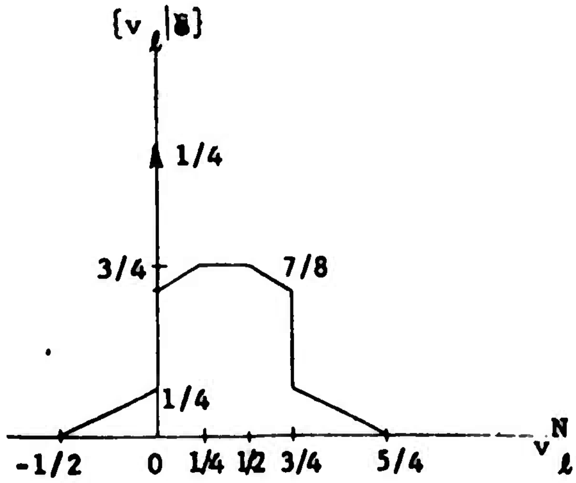


Figure A.1e.

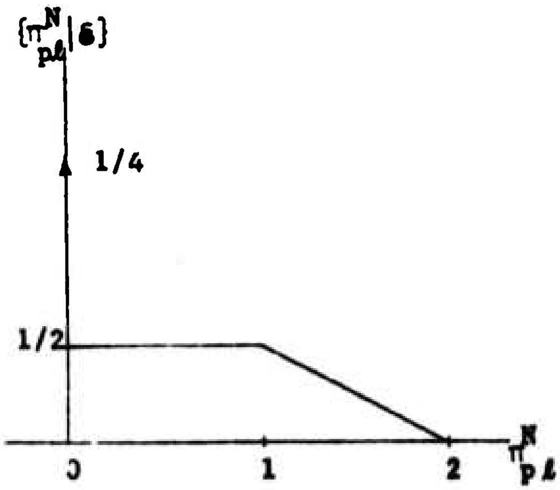


Figure A.1f.

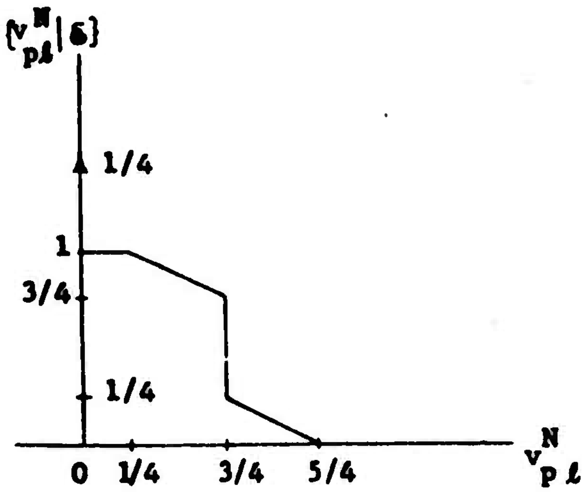


Figure A.1g.

Figure A.1(a-g). Probability density functions for profit and value of information (continued)

perfect information about any of the state variables.

It is possible to derive distributions for the expected profit and the value of perfect information when the information is obtained sequentially. The resulting distributions depend on the prices of the observables in a rather complicated manner. To illustrate this fact, consider the case where we are given perfect information about p and then have the option of paying K_ℓ to learn ℓ. The previous calculations showed that the value of perfect information and the expected profit in this case were both functions of K_ℓ. The expected profit, when we are given perfect information about p, is given by

$$\bar{\pi}_p = \underset{p}{E} \max \begin{Bmatrix} \underset{\ell}{E} \max\limits_b \pi(p,\ell,b) - K_\ell \\ \\ \max\limits_b \underset{\ell}{E} \pi(p,\ell,b) \end{Bmatrix}$$

$$= \begin{Bmatrix} 56/96 - K_\ell & : K_\ell \le 1/8 \\ 60/96 + (4/3)K_\ell\sqrt{2K_\ell} - 2K_\ell & : 1/8 < K_\ell < 1/2 \\ 28/96 & : 1/2 \le K_\ell \end{Bmatrix}$$

To find the probability density function for π_p, we must first determine our optimum bid and the profit for every possible value of p and ℓ. When we have $\pi_p(p,\ell)$, we can find $\{\pi_p|\mathcal{E}\}$, the density function for π_p.

In our previous calculations we found that our decision to pay K_ℓ to learn ℓ depends on the values of p and K_ℓ. There were three cases:

(1) If $K_\ell \le 1/8$, we should always pay K_ℓ to learn ℓ regardless of the value of p. In this case the optimum bid is

$$b = \begin{cases} \bar{\ell} - p : \ell > p \\ 0 \quad : \ell \le p \end{cases}$$

The resulting profit as a function of p , ℓ , and K_ℓ is

$$\pi_p = \begin{cases} \ell - p - K_\ell : \ell > p \\ -K_\ell \quad : \ell \le p \end{cases}$$

This profit function has exactly the same form as the profit function that results from being given perfect information about both p and ℓ simultaneously, except that we must pay K_ℓ in the sequential case. Therefore the probability density function for π_p is just the distribution in Fig. A.1f for the profit when we are given perfect information about both p and ℓ , shifted to the left by K_ℓ . This distribution is shown in Fig. A.2, and it is given by

$$\{\pi_p|\&\} = (1/4)\ \delta(\pi_p + K_\ell) + (1/2)[u(\pi_p + K_\ell) - u(\pi_p - (1-K_\ell))]$$
$$+ [1 - (\pi_p + K_\ell)/2]\ [u(\pi_p - (1-K_\ell)) - u(\pi_p - (2-K_\ell))]$$

($\delta(.)$ is the unit impulse function. $u(.)$ is the unit step function)

(2) If $1/8 < K_\ell < 1/2$, we should pay K_ℓ to learn ℓ when $0 \le p \le 2(1 - \sqrt{2K_\ell})$. Otherwise we should not pay to learn ℓ . In this case the optimum bid is

$$b = \begin{cases} \bar{\ell} \quad : \ell > p\ ,\ 0 \le p \le 2(1 - \sqrt{2K_\ell}) \\ \infty \quad : \ell \le p\ ,\ 0 \le p \le 2(1 - \sqrt{2K_\ell}) \\ 1 + p/2 : 2(1 - \sqrt{2K_\ell}) < p \le 1 \end{cases}$$

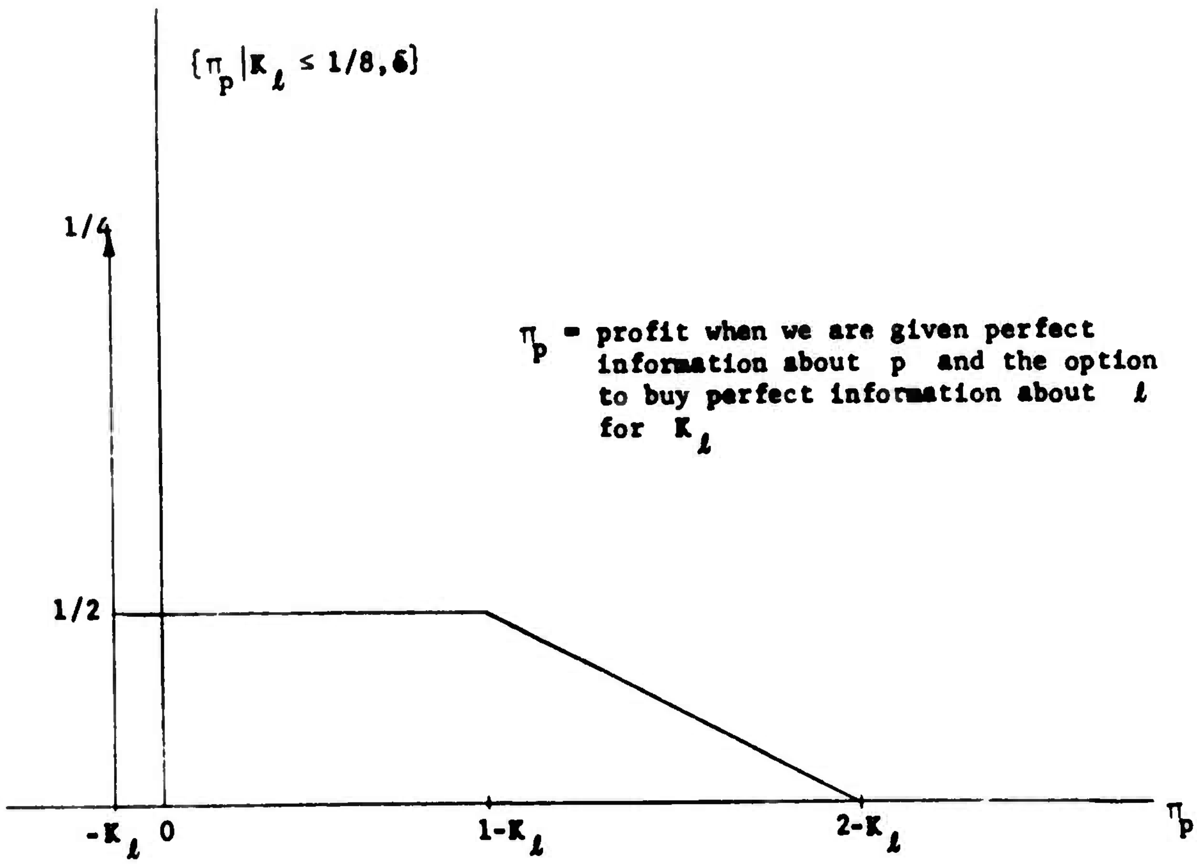


Figure A.2. Probability density function for profit if we learn p with an option to pay K_ℓ for ℓ ($K_\ell \le 1/8$)

The resulting profit as a function of p , ℓ , and K_ℓ is

$$\pi_p = \begin{cases} \bar{\ell} - p - K_\ell : \ell > p \ , \quad 0 \le p \le 2(1 - \sqrt{2K_\ell}) \\ -K_\ell \quad : \ell \le p \ , \quad 0 \le p \le 2(1 - \sqrt{2K_\ell}) \\ 1 - p/2 \quad : 1 + p/2 < \ell \ , \quad 2(1 - \sqrt{2K_\ell}) < p \le 1 \\ 0 \quad : 1 + p/2 \ge \ell \ , \quad 2(1 - \sqrt{2K_\ell}) < p \le 1 \end{cases}$$

This function is shown in Fig. A.3a. Since we know the profit as a function of p and ℓ , and since we know the joint probability density function for p and ℓ , we can carry out the calculations necessary to find the probability density function for π_p [3,14]. The resulting function is

$$\begin{aligned} \{\pi_p|\mathcal{E}\} = & (2\sqrt{2K_\ell} - K_\ell - 7/8)\ \delta(\pi_p) + (1 - 2\sqrt{2K_\ell} + 2K_\ell)\ \delta(\pi_p + K_\ell) \\ & + (1 - \sqrt{2K_\ell})\ [u(\pi_p + K_\ell) - u(\pi_p - (2\sqrt{2K_\ell} - K_\ell))] \\ & + (1 - (\pi_p + K_\ell)/2)\ [u(\pi_p - (2\sqrt{2K_\ell} - K_\ell) - u(\pi_p - (2-K_\ell))] \\ & + \pi_p[u(\pi_p - 1/2) - u(\pi_p - \sqrt{2K_\ell})] \end{aligned}$$

This function is shown in Fig. A-3b.

(3) If $K_\ell \ge 1/2$, we will never pay to learn ℓ , regardless of the value of p . In this case the optimum bid is

$$b = 1 + p/2$$

The resulting profit as a function of p and ℓ is

$$\pi_p = \begin{cases} 1 - p/2 : 1 + p/2 < \ell \\ 0 \quad : 1 + p/2 \ge \ell \end{cases}$$

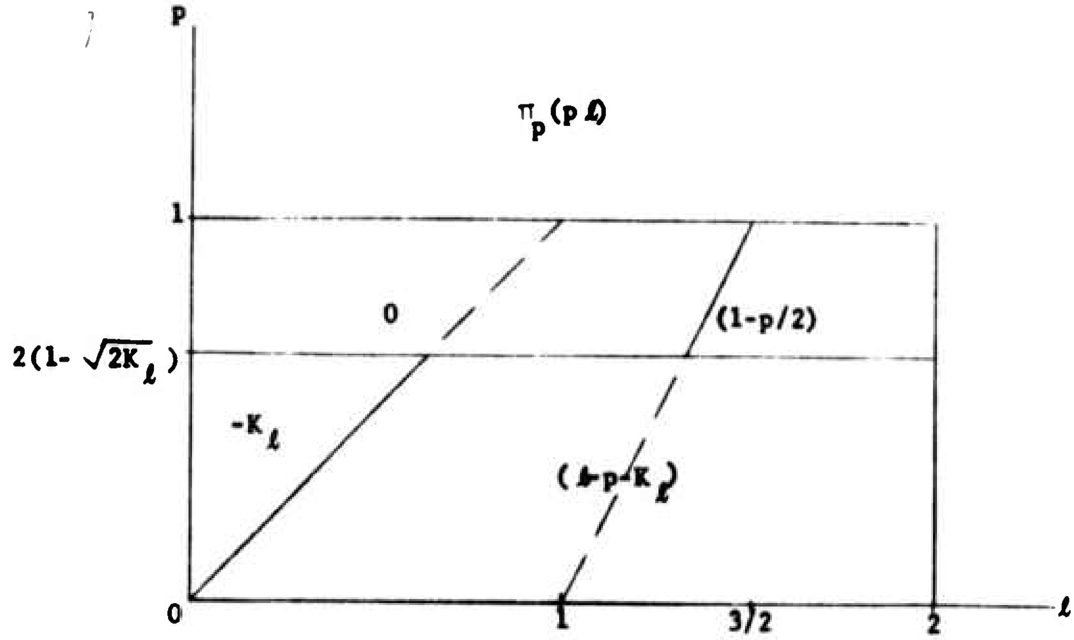


Figure A.3a. Profit when we learn p with an option to pay K_ℓ for ℓ $(1/8 < K_\ell < 1/2)$

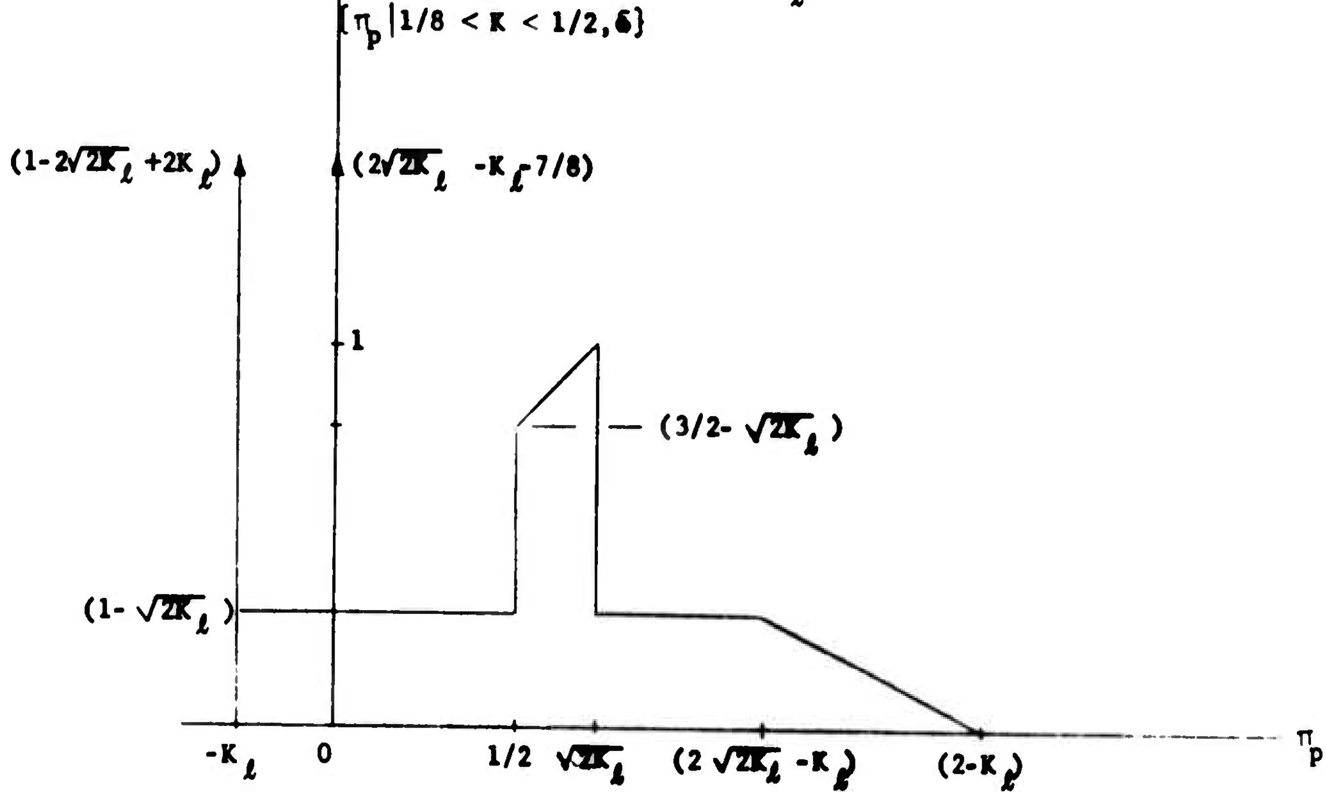


Figure A.3b. Probability density function for profit when we learn p with an option to pay K_ℓ for ℓ $(1/8 < K_\ell < 1/2)$

This is just the profit that we would achieve by learning the actual value of p only. Thus the resulting probability density function is the distribution in Fig. A.1b for the profit from learning p alone. This distribution is shown in Fig. A.4, and it is given by

$$\{\pi_p | \mathcal{E}\} = (5/8)\ \delta(\pi_p) + \pi_p [u(\pi_p - 1/2) - u(\pi_p - 1)]$$

Determining the expected value of π_p with the probability density functions in Figs. A.2, A.3, and A.4 yields the expressions for the expected profit that we found previously.

It is possible to use the same procedure to find the probability density function for the value of perfect information about p, assuming that perfect information about ℓ can be purchased for K_ℓ. It is also possible to find the probability density functions for the value of information and the expected profit when we are given perfect information about ℓ and then offered perfect information about p for K_p. However since these calculations involve no new concepts, they are not included here.

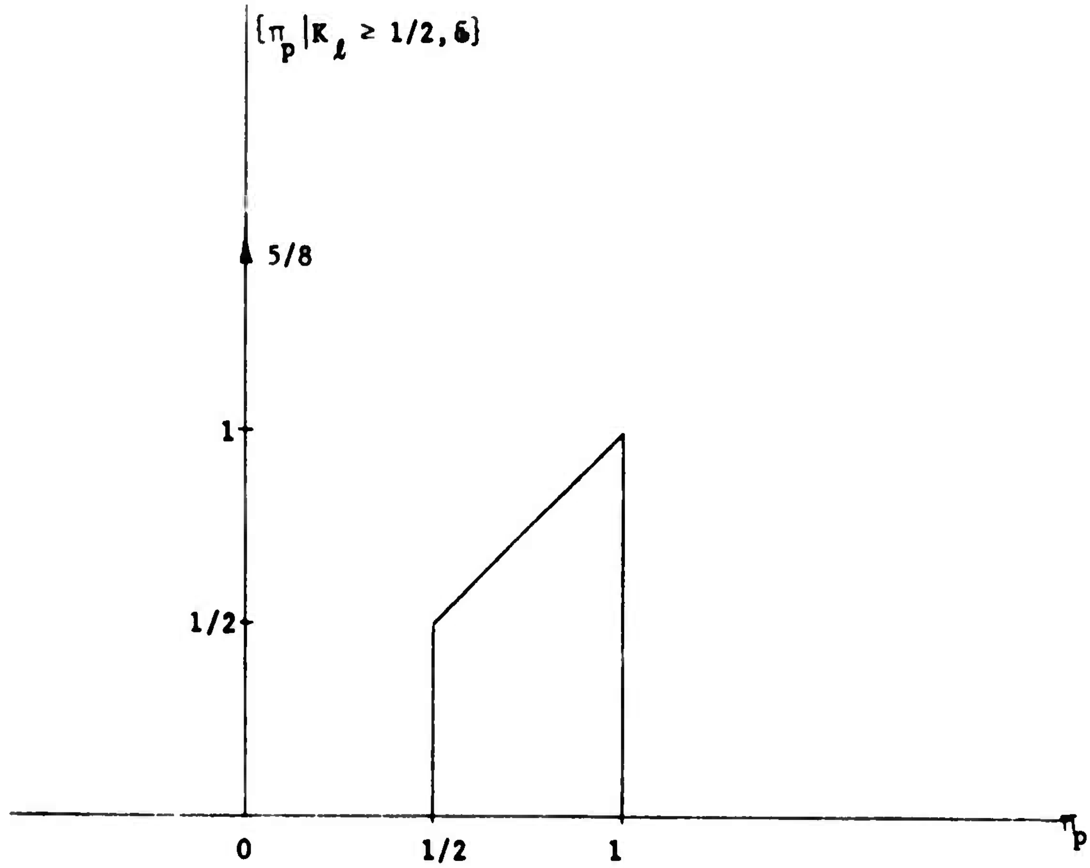


Figure A.4. Probability density function for profit when we learn p with an option to pay K_ℓ for ℓ $(1/2 \leq K_\ell)$

APPENDIX B

THE WEATHER FORECASTING PROBLEM IN DETAIL

In Chapter 2 we formulated a decision problem where the observables represent imperfect information. The problem is to predict the weather, which is described by the state variable x . The probability mass function for x is shown in Fig. 3.1. Having forecast the weather, we can decide to set up an activity indoors or outdoors. This decision is represented by the control variable c . Setting c equal to zero is equivalent to setting up the activity outdoors, and setting c equal to one is equivalent to setting up the activity indoors. The profit function is

$$\pi(x,c) = \left\{ \begin{array}{l} 1 : c = x \\ \\ 0 : \text{otherwise} \end{array} \right\}$$

Two weather forecasts are available, and they can give us imperfect information about the state of the weather. The weather forecasts are described by two observables, y_1 and y_2 . If y_i equals one, the i^{th} forecast is rain, and if y_i equals zero, the i^{th} forecast is fair weather $(i=1,2)$. Our state of information about i^{th} forecast is given by the conditional probability mass function in Fig. 3.2.

Since we have decided that knowing either forecast would not tell us anything about the other forecast when we know the actual state of the weather, y_1 and y_2 must be conditionally independent.

$$\{y_1, y_2 | x, \mathcal{E}\} = \{y_1 | x, \mathcal{E}\}\ \{y_2 | x, \mathcal{E}\}$$

$$= \begin{cases} 36/100 : x = y_1 = y_2 \ , \quad x = 0,1 \\ 24/100 : x = 1 - y_1 = y_2 \ , \quad x = 0,1 \\ 24/100 : x = y_1 = 1 - y_2 \ , \quad x = 0,1 \\ 16/100 : x = 1 - y_1 = 1 - y_2 \ , \quad x = 0,1 \\ 0 \quad : \text{otherwise} \end{cases}$$

In the following discussion we will also need to know several other probability mass functions that can be derived from the information above [14]. These mass functions can be determined easily from the probability tree shown in Fig. B.1.

$$\{y_i | \mathcal{E}\} = \sum_{x=0}^{1} \{y_i | x, \mathcal{E}\}\ \{x | \mathcal{E}\} = \begin{cases} 7/15 : y_i = 0 \\ 8/15 : y_i = 1 \\ 0 \quad : \text{otherwise} \end{cases} \quad (i = 1,2)$$

$$\{y_1, y_2 | \mathcal{E}\} = \sum_{x=0}^{1} \{y_1, y_2 | x, \mathcal{E}\}\ \{x | \mathcal{E}\}$$

$$= \begin{cases} 68/300 : y_1 = 0 \ , \quad y_2 = 0 \\ 72/300 : y_1 = 0 \ , \quad y_2 = 1 \\ 72/300 : y_1 = 1 \ , \quad y_2 = 0 \\ 88/300 : y_1 = 1 \ , \quad y_2 = 1 \\ 0 \quad : \text{otherwise} \end{cases}$$

$$\{y_i | y_j, \mathcal{E}\} = \frac{\{y_1, y_2 | \mathcal{E}\}}{\{y_j | \mathcal{E}\}} = \begin{cases} 17/35 : y_i = 0 \ , \quad y_j = 0 \\ 9/20 : y_i = 0 \ , \quad y_j = 1 \\ 18/35 : y_i = 1 \ , \quad y_j = 0 \\ 11/20 : y_i = 1 \ , \quad y_j = 1 \\ 0 \quad : \text{otherwise} \end{cases} \quad \begin{bmatrix} i,j = 1,2 \\ i \neq j \end{bmatrix}$$

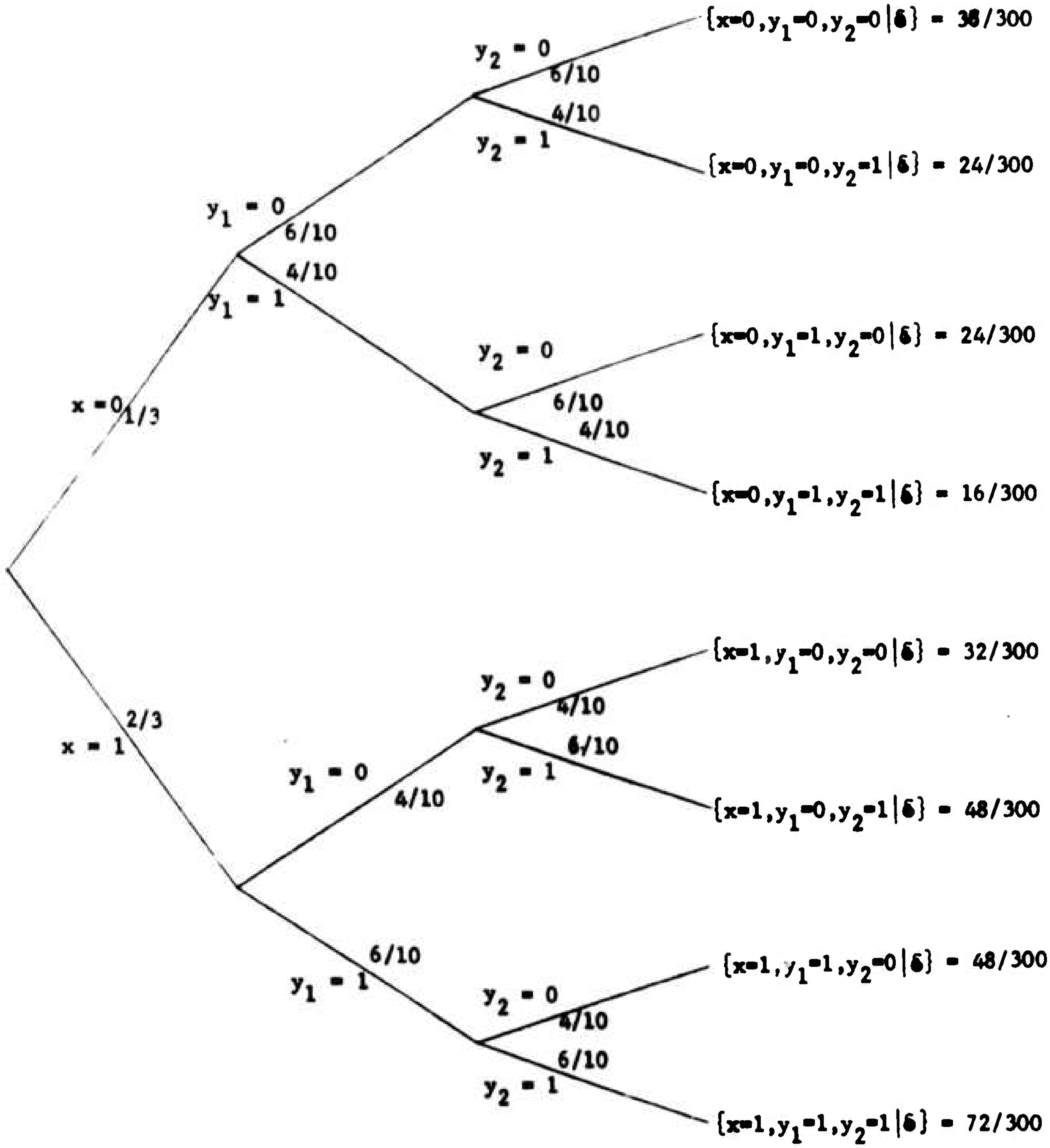


Figure B.1. Probability tree for weather-forecasting problem

$$\{x|y_i,\mathcal{E}\} = \frac{\{y_i|x,\mathcal{E}\}\ \{x|\mathcal{E}\}}{\{y_i|\mathcal{E}\}} = \begin{cases} 3/7 : x = 0\ , & y_i = 0 \\ 4/7 : x = 1\ , & y_i = 0 \\ 1/4 : x = 1\ , & y_i = 1 \\ 3/4 : x = 1\ , & y_i = 1 \\ 0\ : \text{otherwise} \end{cases} \quad (i = 1,2)$$

$$\{x|y_1,y_2,\mathcal{E}\} = \frac{\{y_1,y_2|x,\mathcal{E}\}\ \{x|\mathcal{E}\}}{\{y_1,y_2|\mathcal{E}\}} = \begin{cases} 9/17 : x = 0\ , & y_1 = 0\ , & y_2 = 0 \\ 8/17 : x = 1\ , & y_1 = 0\ , & y_2 = 0 \\ 1/3 : x = 0\ , & y_1 = 0\ , & y_2 = 1 \\ 2/3 : x = 1\ , & y_1 = 0\ , & y_2 = 1 \\ 1/3 : x = 0\ , & y_1 = 1\ , & y_2 = 0 \\ 2/3 : x = 1\ , & y_1 = 1\ , & y_2 = 0 \\ 2/11 : x = 0\ , & y_1 = 1\ , & y_2 = 1 \\ 9/11 : x = 1\ , & y_1 = 1\ , & y_2 = 1 \\ 0\ : \text{otherwise} \end{cases}$$

We can use this information to determine the value of sequential imperfect information. If we decide not to pay for either forecast, our expected profit is

$$\max_c \mathop{E}_x \pi(x,c) = \max_c \mathop{E}_x \begin{Bmatrix} 1 : c = x \\ \\ 0 : \text{otherwise} \end{Bmatrix} = 2/3$$

The maximum profit occurs when we set c equal to one, or, in other words, decide to set up the activity indoors.

The Values of Individual and Simultaneous Information

If we are given the first weather forecasting company's prediction

free of charge and do not intend to buy the second forecast, then our expected profit after learning this information is

$$\underset{y_1}{D} \max_{c} \underset{x}{E}\, \pi(x,c) = \begin{cases} 4/7 : y_1 = 0 \\ 3/4 : y_1 = 1 \\ 0 \quad : \text{otherwise} \end{cases}$$

The maximum expected profit occurs when $c = 1$, or when we set up the activity indoors, regardless of y_1 . Before we are given y_1 our expected profit is

$$\underset{y_1}{E} \max_{c} \underset{x}{E}\, \pi(x,c) = \underset{y_1}{E} \left\{ \begin{array}{l} 4/7 : y_1 = 0 \\ 3/4 : y_1 = 1 \\ 0 \quad : \text{otherwise} \end{array} \right\} = 2/3$$

Using this result we find that the value of learning the first company's forecast by itself is

$$V^N_{y_1} = (\underset{y_1}{E} \max_{c} \underset{x}{E} - \max_{c} \underset{x}{E})\, \pi(x,c) = 2/3 - 2/3 = 0$$

The fact that the value of individual information about y_1 is zero should not surprise us. Before we know y_1 our best decision is to set up the activity indoors, and after we learn y_1 this is still the best decision. Since the information cannot change our decision, the value of the information is zero.

Exactly the same reasoning and calculations show that the value of individual information about y_2 , the second company's forecast, is also zero.

$$V^N_{y_2} = (\underset{y_2}{E} \max_{c} \underset{x}{E} - \max_{c} \underset{x}{E})\, \pi(x,c) = 0$$

As before, our best decision is to set up the activity indoors regardless of what the second weather forecasting company thinks the weather is going to be.

If we are given both weather forecasts simultaneously, our expected profit after learning the forecasts is

$$\underset{y_1}{D}\ \underset{y_2}{D}\ \max_{c}\ \underset{x}{E}\ \pi(x,c) = \begin{Bmatrix} 9/17 : y_1 = 0\ ,\ \ y_2 = 0 \\ 2/3 : y_1 = 0\ ,\ \ y_2 = 1 \\ 2/3 : y_1 = 1\ ,\ \ y_2 = 0 \\ 9/11 : y_1 = 1\ ,\ \ y_2 = 1 \\ 0\ \ : \text{otherwise} \end{Bmatrix}$$

The maximum expected profit occurs when

$$c = \begin{Bmatrix} 0 : y_1 = 0\ ,\ \ y_2 = 0 \\ 1 : \text{otherwise} \end{Bmatrix}$$

In other words we should set up outdoors if both companies forecast fair weather, and stay indoors if either one forecasts rain. This time the information can cause us to change our decision so the value of learning y_1 and y_2 simultaneously will be positive.

Before we are given the two weather forecasts, our expected profit is

$$\underset{y_1}{E}\ \underset{y_2}{E}\ \max_{c}\ \underset{x}{E}\ \pi(x,c) = \underset{y_1}{E}\ \underset{y_2}{E} \begin{Bmatrix} 9/17 : y_1 = 0\ ,\ \ y_2 = 0 \\ 2/3 : y_1 = 0\ ,\ \ y_2 = 1 \\ 2/3 : y_1 = 1\ ,\ \ y_2 = 0 \\ 9/11 : y_1 = 1\ ,\ \ y_2 = 1 \\ 0\ \ : \text{otherwise} \end{Bmatrix} = 51/75$$

Using this result we find that the value of learning both weather forecasts simultaneously is

$$V^N_{y_1 y_2} = (\underset{y_1}{E}\ \underset{y_2}{E}\ \underset{c}{\max}\ \underset{x}{E} - \underset{c}{\max}\ \underset{x}{E})\ \pi(x,c) = 51/75 - 2/3 = 1/75$$

Comparing $V^N_{y_1 y_2}$ with $V^N_{y_1}$ amd $V^N_{y_2}$, we see that the value of learning both forecasts simultaneously does not equal the sum of the values of individual information about the two forecasts. In general $V^N_{y_1 y_2}$ can be greater or less than the sum of $V^N_{y_1}$ and $V^N_{y_2}$, in exactly the same way that the value of learning two pieces of perfect information can be greater or less than the sum of the values of individual perfect information.

Assume that the two weather forecasts y_1 and y_2 can be purchased for prices K_{y_1} and K_{y_2}, respectively. We are willing to pay for the i[th] weather forecast when $K_{y_i} < V^N_{y_i}$ $(i = 1,2)$. We are willing to pay for both forecasts simultaneously when $(K_{y_1} + K_{y_2}) < V^N_{y_1 y_2}$. These decision rules are shown graphically in Fig. B.2. Since $V^N_{y_1}$ and $V^N_{y_2}$ are zero, we will not pay to learn just one of the two forecasts. However learning both forecasts might change our decision to set up the activity indoors, so there are pairs of prices such that we are willing to learn y_1 and y_2 simultaneously. These pairs of prices are represented by points below and to the left of the line A-B in Fig. B.2.

We can use $V^N_{y_1 y_2}$ to define two related quantities $V^R_{y_1}(K_{y_2})$ and $V^R_{y_2}(K_{y_1})$. $V^R_{y_1}(K_{y_2})$ is the residual value of learning y_1 when we must buy the two weather forecasts simultaneously. We define

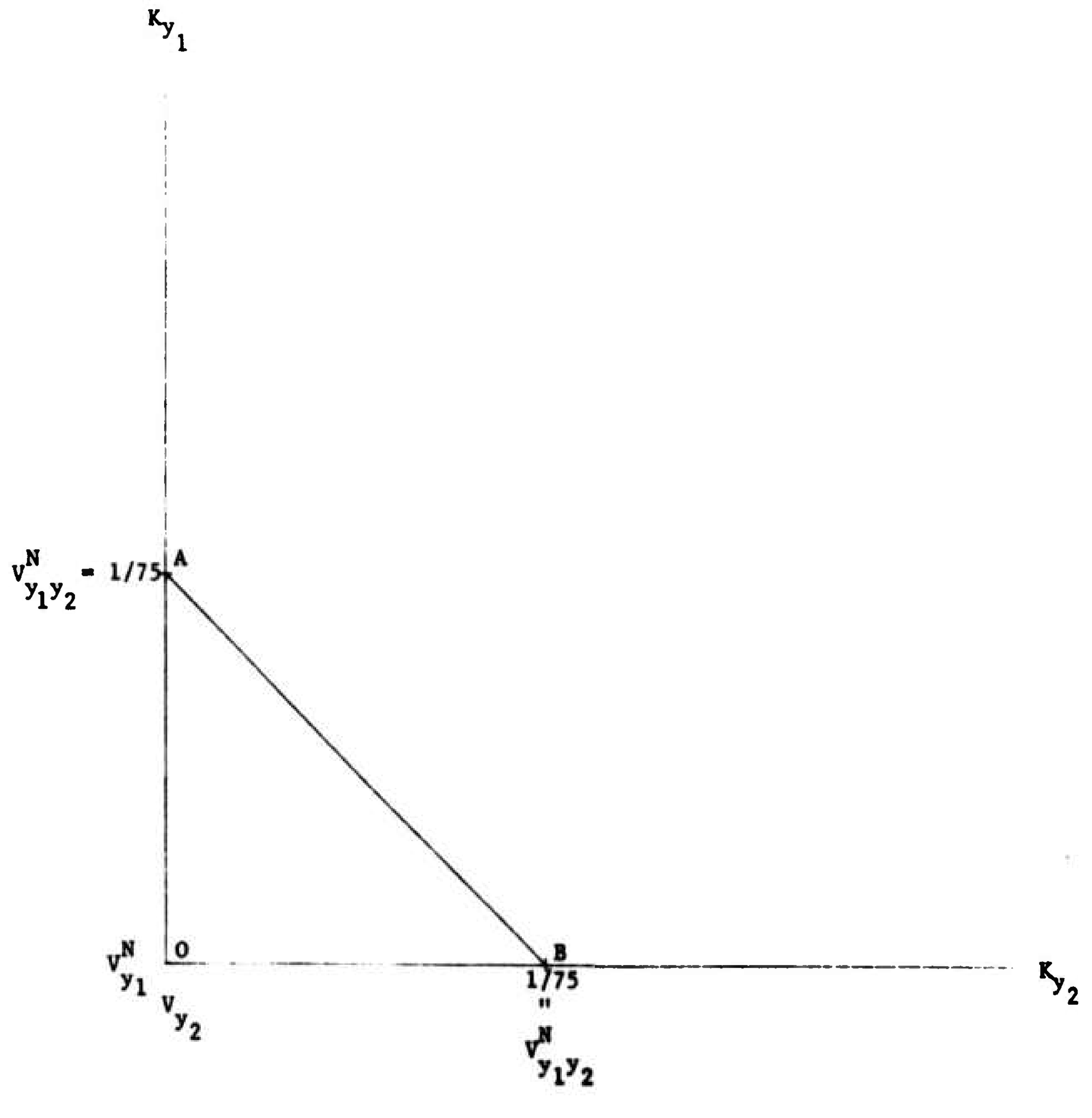


Figure B.2. Price diagram with individual and simultaneous information

$V^R_{y_1}(K_{y_2})$ as follows:

$$V^R_{y_1}(K_{y_2}) = V^N_{y_1 y_2} - K_{y_2} = 1/75 - K_{y_2}$$

We are willing to buy both pieces of information simultaneously when

$$K_{y_1} < V^R_{y_1}(K_{y_2})$$

This is the same decision rule as the one shown in Fig. B.2. In exactly the same way, the residual value of information about y_2 is defined to be

$$V^R_{y_2}(K_{y_1}) = V^N_{y_1 y_2} - K_{y_1} = 1/75 - K_{y_1}$$

We are willing to buy both pieces of information simultaneously when

$$K_{y_2} < V^R_{y_2}(K_{y_1})$$

The Value of Sequential Information

Now consider the situation where the two weather forecasts can be purchased sequentially. Assume for a moment that we already know the first forecast and we are trying to decide whether or not to pay for the second. How much is the second piece of information worth to us?

If we decide not to learn the second forecast, then our expected profit as a function of y_1 is

$$\underset{y_1}{D} \max_c \underset{x}{E}\, \pi(x,c) = \begin{Bmatrix} 4/7 : y_1 = 0 \\ 3/4 : y_1 = 1 \\ 0 \;\; : \text{otherwise} \end{Bmatrix}$$

The maximum expected profit occurs when $c = 1$. On the other hand, if

we decide to learn the second weather forecast, the expected profit as a function of y_1 is

$$\underset{y_1}{D}\ \underset{y_2}{E}\ \max_c\ \underset{x}{E}\ \pi(x,c) = \begin{cases} 3/5 : y_1 = 0 \\ 3/4 : y_1 = 1 \\ 0 \ : \text{otherwise} \end{cases}$$

The maximum profit occurs when

$$c = \begin{cases} 0 : y_1 = 0\ ,\ \ y_2 = 0 \\ 1 : \text{otherwise} \end{cases}$$

The increase in expected profit caused by learning y_2 , when we already know y_1 , is the difference between the two expressions above.

$$\underset{y_1}{D}\ (\underset{y_2}{E}\ \max_c\ \underset{x}{E} - \max_c\ \underset{x}{E})\ \pi(x,c) = \begin{cases} 1/35 : y_1 = 0 \\ 0 \quad : \text{otherwise} \end{cases}$$

Now suppose that the price of the second company's weather forecast is K_{y_2} . We should pay this price for the forecast when the increase in expected profit exceeds K_{y_2} ; otherwise we should not buy the information. Thus we should pay for the second forecast when

$$K_{y_2} < \underset{y_1}{D}\ (\underset{y_2}{E}\ \max_c\ \underset{x}{E} - \max_c\ \underset{x}{E})\ \pi(x,c)$$

This decision rule is shown in Fig. B.3. It is obvious from the inequality, or from Fig. B.3, that for certain values of K_{y_2} the decision to buy or not buy information about y_2 depends on y_1 , the forecast we learned earlier. There are two cases:

(1) If $K_{y_2} < 1/35$, we may or may not pay K_{y_2} to learn y_2 ,

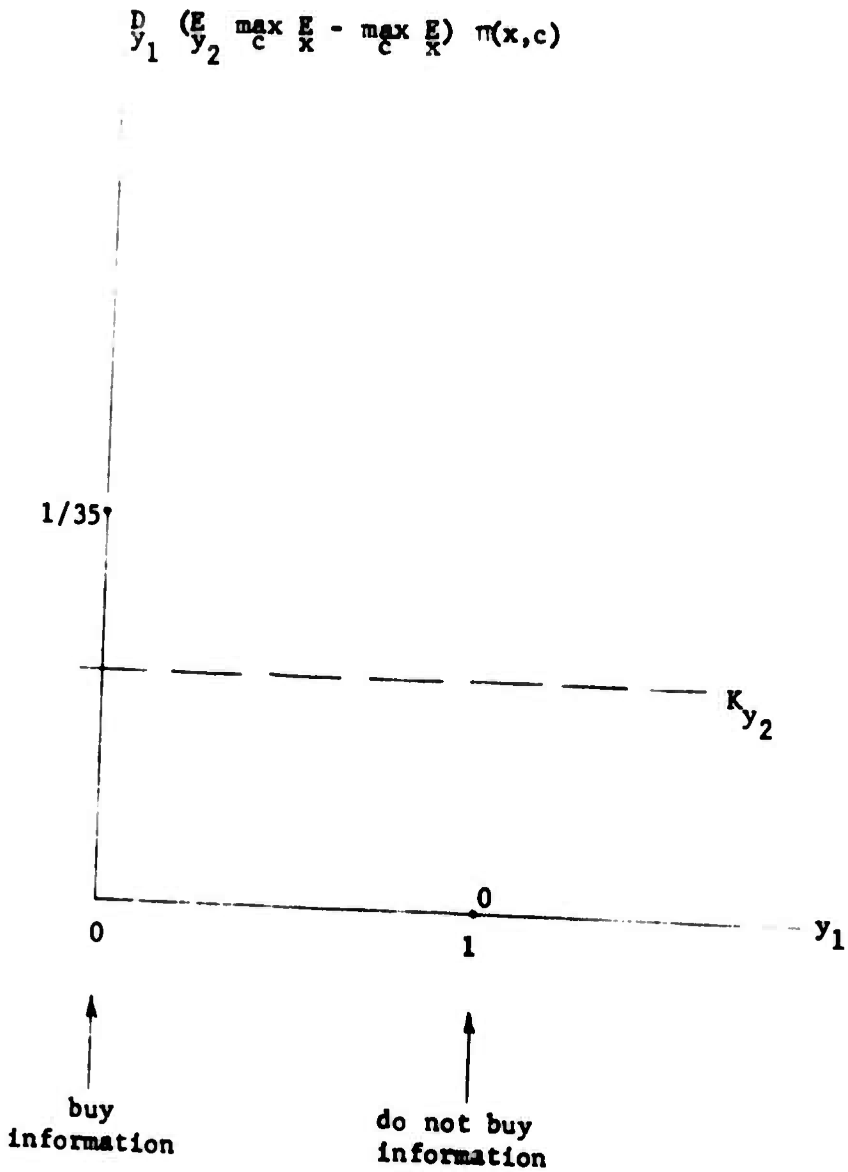


Figure B.3. Increase in expected profit caused by learning y_2 when y_1 is known

depending on the value of y_1 . We will pay to learn y_2 when

$$K_{y_2} < \begin{cases} 1/35 : y_1 = 0 \\ 0 \quad : \text{otherwise} \end{cases}$$

or, in other words, when $y_1 = 0$. Thus the expected profit as a function of y_1 is

$$\underset{y_1}{D} \max \begin{Bmatrix} \underset{y_2}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_2} \\ \max_c \underset{x}{E} \pi(x,c) \end{Bmatrix}$$

$$= \begin{Bmatrix} \underset{y_1}{D} \underset{y_2}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_2} : y_1 = 0 \\ \underset{y_1}{D} \max_c \underset{x}{E} \pi(x,c) \quad : y_1 = 1 \end{Bmatrix}$$

$$= \begin{Bmatrix} 3/5 - K_{y_2} : y_1 = 0 \\ 3/4 \quad : y_1 = 1 \end{Bmatrix}$$

(2) If $K_{y_2} \geq 1/35$, we will never pay K_{y_2} to learn y_2 , regardless of the value of y_1 , because K_{y_2} will always exceed the increase in the expected profit. In this case, the expected profit as a function of y_1 is

$$\underset{y_1}{D} \max_c \underset{x}{E} \pi(x,c) = \begin{Bmatrix} 4/7 : y_1 = 0 \\ 3/4 : y_1 = 1 \end{Bmatrix}$$

These equations give us the expected profit as a function of y_1 and K_{y_2} . To find the expected profit before we learn y_1 we need to take the expected value with respect to y_1 of the previous results.

Thus our expected profit when we learn that we will be given the first weather forecast, but before we receive the information, is

$$\mathop{E}_{y_1} \max \left\{ \begin{array}{l} \mathop{E}_{y_2} \max_c \mathop{E}_x (x,c) - K_{y_2} \\ \\ \max_c \mathop{E}_x (x,c) \end{array} \right\} = \left\{ \begin{array}{ll} 51/75 - (7/15)K_{y_2} & : K_{y_2} < 1/35 \\ \\ 2/3 & : K_{y_2} \geq 1/35 \end{array} \right\}$$

Since our expected profit without any information is $\max_c \mathop{E}_x \pi(x,c)$, we must subtract this quantity from the result above to get the value of sequential information about the first weather forecast $V_{y_1}(K_{y_2})$.

$$V_{y_1}(K_{y_2}) = \mathop{E}_{y_1} \max \left\{ \begin{array}{l} \mathop{E}_{y_2} \max_c \mathop{E}_x \pi(x,c) - K_{y_2} \\ \\ \max_c \mathop{E}_x \pi(x,c) \end{array} \right\} - \max_c \mathop{E}_x \pi(x,c)$$

$$= \left\{ \begin{array}{ll} 1/75 - (7/15)K_{y_2} & : K < 1/35 \\ \\ 0 & : K_{y_2} \geq 1/35 \end{array} \right\}$$

If we are offered the first weather forecast for K_{y_1} , we will accept the offer when $K_{y_1} < V_{y_1}(K_{y_2})$. However V_{y_1} depends on K_{y_2} , so we must know both K_{y_1} and K_{y_2} before we can decide whether or not to pay for the first forecast. In terms of K_{y_1} and K_{y_2} , we should pay K_{y_1} for the first weather forecast when

$$K_{y_1} < \left\{ \begin{array}{ll} 1/75 - (7/15)K_{y_2} & : K_{y_2} < 1/35 \\ \\ 0 & : K_{y_2} \geq 1/35 \end{array} \right\}$$

This decision rule is shown graphically in Fig. 3.3. We are willing to pay K_{y_1} to learn the first weather forecast for any pair of prices represented by a point below the line A-E-D.

Since our initial state of information is completely symmetrical with respect to the two weather forecasting companies, the calculation to determine the value of sequential information about the second weather forecast y_2 looks like the calculation above with the subscripts changed. We start by assuming that we know the second forecast, and ask how much it would be worth to also learn the first forecast. Following the same reasoning as before, we find that the value of sequential information about the second forecast is

$$V_{y_2}(K_{y_1}) = \underset{y_2}{E} \max \left\{ \begin{array}{l} \underset{y_1}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_1} \\ \\ \max E \pi(x,c) \end{array} \right\} - \max_c \underset{x}{E} \pi(x,c)$$

$$= \left\{ \begin{array}{ll} 1/75 - (7/15)K_{y_1} & : K_{y_1} < 1/35 \\ \\ 0 & : K_{y_1} \geq 1/35 \end{array} \right\}$$

We are willing to pay K_{y_2} to learn the second forecast when $K_{y_2} < V_{y_2}(K_{y_1})$. This decision rule is shown in Fig. 3.3. We are willing to pay K_{y_2} to learn y_2 for any pair of prices represented by a point to the left of the line B-E-C.

The Relative Values of $V^N_{y_i}$, $V^R_{y_i}(K_{y_j})$, and $V_{y_i}(K_{y_j})$

Thus we should be willing to pay for one of the two pieces of imperfect information if we are offered any pair of prices represented by a point below or to the left of the boundary C-E-D in Fig. 3.3. This set of price pairs, which we call the set of feasible prices, includes all of the feasible pairs of prices that we found previously by considering individual and simultaneous information. Using the equations

derived above, we can show that this must always be the case. If we only buy one of the two forecasts by itself, then the value of the information is

$$V^N_{y_i} = \underset{y_i}{E} \max_c \underset{x}{E} \pi(x,c) - \max_c \underset{x}{E} \pi(x,c) \qquad (i = 1,2)$$

If we buy both forecasts simultaneously, then the value of the information is

$$V^R_{y_i}(K_{y_j}) = \underset{y_i}{E} \left(\underset{y_j}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_j}\right) - \max_c \underset{x}{E} \pi(x,c) \qquad \begin{matrix}(i,j = 1,2) \\ (i \neq j)\end{matrix}$$

If we buy forecasts sequentially the value of the information is

$$V_{y_i}(K_{y_j}) = \underset{y_i}{E} \max \left\{ \begin{matrix} \underset{y_j}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_j} \\ \\ \max_c \underset{x}{E} \pi(x,c) \end{matrix} \right\} - \max E \pi(x,c) \qquad \begin{matrix}(i,j = 1,2) \\ (i \neq j)\end{matrix}$$

From the form of these equations we can show that $V_{y_i}(K_{y_j})$ must be at least as large as $V^N_{y_i}$ or $V^R_{y_i}(K_{y_j})$ for any value of K_{y_j} and for $i,j = 1,2$; $i \neq j$. Obviously, for any value of K_{y_j} :

$$\max \left\{ \begin{matrix} \underset{y_i}{D} \underset{y_j}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_j} \\ \\ \underset{y_i}{D} \max E \pi(x,c) \end{matrix} \right\} \geq \underset{y_i}{D} \underset{y_j}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_j}$$

$$\max \left\{ \begin{matrix} \underset{y_i}{D} \underset{y_j}{E} \max_c \underset{x}{E} \pi(x,c) - K_{y_j} \\ \\ \underset{y_i}{D} \max_c \underset{x}{E} \pi(x,c) \end{matrix} \right\} \geq \underset{y_i}{D} \max_c \underset{x}{E} \pi(x,c) \qquad \begin{matrix}(i,j = 1,2) \\ (i \neq j)\end{matrix}$$

Taking the expected value with respect to y_i and subtracting

$\max_c \mathop{E}_x \pi(x,c)$ from both sides of the inequalities yield:

$$V_{y_i}(K_{y_j}) \geq V^S_{y_i}(K_{y_j}) \qquad (i,j = 1,2)$$

$$V_{y_i}(K_{y_j}) \geq V^N_{y_i} \qquad (i \neq j)$$

Decision Regions in the Price Diagram

Figure 3.3 shows the pairs of prices for which we are willing to pay for one of the forecasts, but it does not show which forecast to buy first. If we decide to buy a forecast, we would expect the best one to buy first to be the one that is least expensive, due to the symmetry of the problem. A quick calculation shows that our intuition is correct. Since we are trying to maximize our expected profit, and since the increase in expected profit is the difference between the value and the cost of the information, we should pay for y_1 first, when

$$V_{y_1}(K_{y_2}) - K_{y_1} > V_{y_2}(K_{y_1}) - K_{y_2}$$

$$\mathop{E}_{y_1} \max \left\{ \begin{array}{l} \mathop{E}_{y_2} \max_c \mathop{E}_x \pi(x,c) - K_{y_2} \\ \\ \max_c \mathop{E}_x \pi(x,c) \end{array} \right\} - \max_x \mathop{E}_c \pi(x,c) - K_{y_1}$$

$$> \mathop{E}_{y_2} \max \left\{ \begin{array}{l} E \max E \pi(x,c) - K \\ \\ \max_c \mathop{E}_x \pi(x,c) \end{array} \right\} - \max_c \mathop{E}_x \pi(x,c) - K_{y_2}$$

For the weather forecasting problem, this becomes

$$\begin{Bmatrix} 1/75 - (7/15)K_{y_2} - K_{y_1} & : K_{y_2} < 1/35 \\ - K_{y_1} & : K_{y_2} \geq 1/35 \end{Bmatrix}$$

$$> \begin{Bmatrix} 1/75 - (7/15)K_{y_1} - K_{y_2} & : K_{y_1} < 1/35 \\ - K_{y_2} & : K_{y_1} \geq 1/35 \end{Bmatrix}$$

In addition we will only pay for y_1 when $V_{y_1}(K_{y_2}) > K$. These inequalities define the decision regions shown in Fig. 3.4.

Characterzing the Decision Regions

The expected value of either forecast can vary with the price of the other forecast, as shown in Fig. 3.4. If we consider only the region in Fig. 3.4 where our best initial decision is to buy information about y_1 , we can see that we would be willing to pay significantly more to learn y_1 when we subsequently have the option to buy the second weather forecast. In fact, without the option to buy the second weather forecast, the first forecast is worthless. When we can subsequently purchase the second forecast and when we restrict our attention to those pairs of prices where our best initial decision is to buy information about y_1 , the first forecast can be worth as much as the K_{y_1} component of the point E in Fig. 3.4. Although there are other pairs of prices for which we would be willing to pay more to learn y_1 first than we would at the pair of prices corresponding to the point E, we could increase our expected profit more by first paying for y_2 instead of y_1 . Thus the maximum amount that we would pay to learn y_1 first is the K_{y_1} component of the point E . Exactly the same

reasoning shows that the maximum amount that we would pay to learn y_2 first is the K_{y_2} component of the point E .

One way to characterize the decision regions in Fig. 3.4 is to determine the maximum amount that we would be willing to pay initially for each piece of information. Since we are dealing with imperfect information, we could bound the value of the weather forecasts by finding the expected value of perfect information about the weather. For the weather forecasting problem this is easy to do since the profit depends on only one state variable,

$$V_x^N = (\max_c \mathop{E}_x - \mathop{E}_x \max_c)\ \pi(x,c) = 1 - 2/3 = 1/3$$

However we can see from Fig. 3.4 that this is not a very tight upper bound. (If the profit depends on more than one state variable, we must use the techniques introduced in Chapter 2 to find an upper bound for the expected value of perfect information.)

A better upper bound on the value of imperfect sequential information is given by the maximum amount that we would pay initially for each piece of information, corresponding to the coordinates of the point E in Fig. 3.4. We can extend the results of the preceding chapter to the case of imperfect information and develop an iterative procedure for finding the coordinates of the point E when there are only two observables. Suppose we first determine the value of individual information about y_1 and y_2 . This gives us $V_{y_1}^N$ and $V_{y_2}^N$, which are both zero in the weather forecasting problem. In Fig. 3.4, this pair of values is represented by a point at the origin. Now use $V_{y_1}^N$ and $V_{y_2}^N$ as the prices K_{y_1} and K_{y_2} in the decision tree

branches in Fig. 3-5b. Solving the decision tree for the expected profits associated with branches B_1 and B_2 (the branches where we first learn y_1 and y_2, respectively), and then subtracting the expected profit associated with branch B_0 (the branch where we do not buy any information), yields the values of sequential information about y_1 and y_2 when the cost of information is given by $V^N_{y_1}$ and $V^N_{y_2}$. Figure B.4a shows that the values of information produced by this calculation correspond to the coordinates of the point F.

Now repeat the procedure by using the coordinates of the point F in Fig. B.4a as the prices K_{y_1} and K_{y_2} in the decision tree branches in Fig. 3.5b. Solving for the expected profits associated with branches B_1 and B_2, and then subtracting the expected profit associated with branch B_0, yields two new values of information. Figure B.4b shows that these new values are the coordinates of the point G. The procedure can be continued by using the coordinates of the point G in the decision tree to get another point that is even closer to the point E.

While this procedure will give us a series of points in the price diagram that converge to the point E, it is not necessary to carry out repeated solutions of the decision tree in Fig. 3.5 to find upper bounds for the values of information about y_1 and y_2. The first iteration using $V^N_{y_1}$ and $V^N_{y_2}$ in the decision tree is all that is necessary to determine the upper bounds. Chapter 4 shows that, whenever there are only two observables, the point F must lie above and to the right of the point E if it does not coincide with the point E. Thus the coordinates of the point F must be at least as large as the coordinates of the point E, which in turn must be at least as large

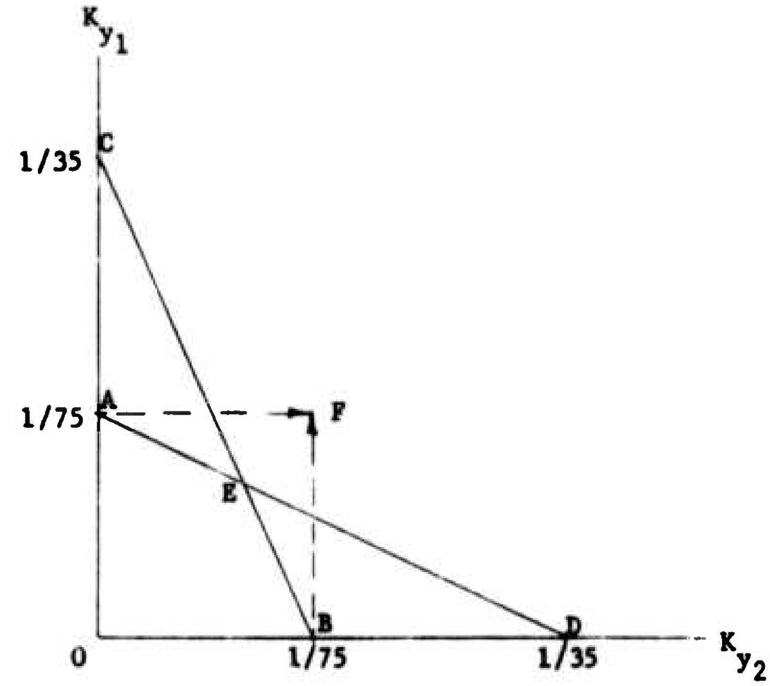


Figure B.4a. Finding the maximum initial prices--first iteration

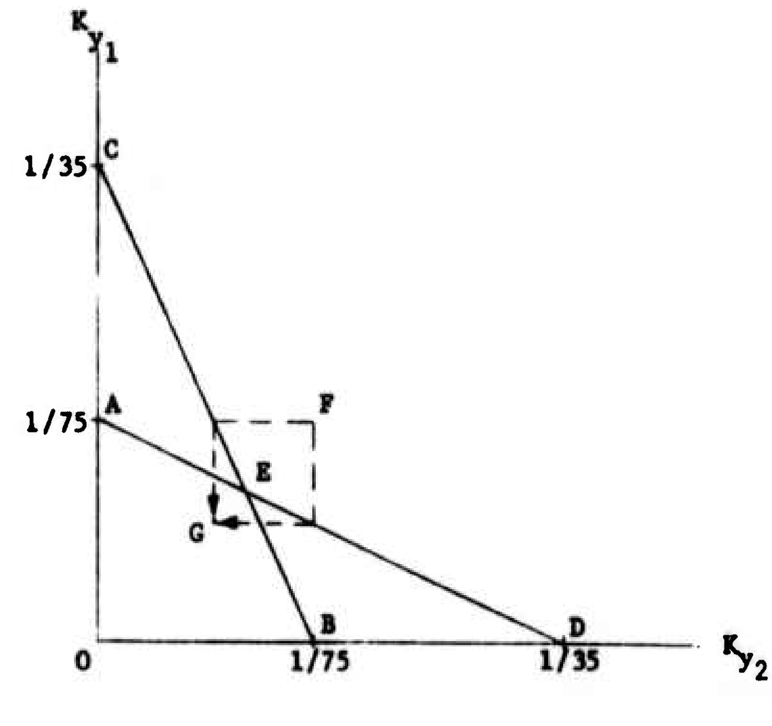


Figure B.4b. Finding the maximum initial prices--second iteration

as the values of information about y_1 and y_2. Therefore the coordinates of the point F are an upper bound for the value of the imperfect weather forecasts.

An interesting feature of the weather forecasting problem is that both of the coordinates of the point F in Fig. B.4 are equal to $V^N_{y_1 y_2}$. This follows from the fact that $V^N_{y_1}$ and $V^N_{y_2}$ are both zero. When we solve the decision tree in Fig. 3.5 with all of the prices set equal to zero, it is obvious that the resulting value of information about any observable will be the value of learning all of the observables simultaneously. No matter which observable we decide to learn first, we will subsequently decide to learn the other observable because it is free. Chapter 4 shows that the value of learning all of the observables simultaneously is always an upper bound for the value of sequential information about any of the observables, even when there are more than two observables.

BIBLIOGRAPHY

[1] Howard, R.A., "Decision Analysis: Applied Decision Theory," Fourth International Conference on Operational Research, Boston, 1966.

[2] Howard, R.A., "Information Value Theory," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-2, No. 1, August 1966.

[3] Howard, R.A., "Value of Information Lotteries," IEEE Transactions on Systems Science and Cybernetics, Vol. SSC-3, June 1967.

[4] Howard, R.A., "The Foundations of Decision Analysis," IEEE Transactions on Systems Science and Cybernetics, Vol. SCC-4, No. 3, September 1968.

[5] Howard, R.A., "Risk Preference," Unpublished paper, Dept. of Engineering-Economic Systems, Stanford University.

[6] IEEE Transactions of Systems Science and Cybernetics, Special Issue on Decision Analysis, Vol. SCC-4, No. 3, September 1968.

[7] LaValle, I.H., "Strategic Situation Theory: A Bayesian Approach to an Individual Player's Choice of Strategies in Noncooperative Games," Unpublished dissertation, Harvard Graduate School of Business Administration, Harvard University, 1966.

[8] LaValle, I.H., "A Bayesian Approach to an Individual Player's Choice of Bid in Competitive Sealed Auctions," Management Science, Vol. 13, No. 7, March 1967.

[9] Luce, R.D., and H. Raiffa, "Games and Decisions: Introduction and Critical Survey," New York: Wiley, 1957.

[10] Pratt, J.W., "Risk Aversion in the Small and in the Large," Econometrica, Vol. 32, 1964.

[11] Pratt, J.W., H. Raiffa, and R.O. Schlaifer, "Introduction to Statistical Decision Theory," McGraw-Hill, 1965.

[12] Raiffa, H., and R.O. Schlaifer, "Applied Statistical Decision Theory," Harvard Graduate School of Business Administration, Harvard University, 1961.

[13] Raiffa, H., "Decision Analysis: Introductory Lectures on Choices under Uncertainty," Addison-Wesley, 1968.

[14] Savage, L.J., "The Foundations of Statistics," New York: John Wiley and Sons, 1954.

[15] Von Neumann, J., and O. Morgenstern, "Theory of Games and Economic Behavior," 2nd Edition, Princeton University Press, 1947.